INTRODUCCIÓN A LA ANAVA

En los temas precedentes hemos estudiado contrastes de hipótesis sobre la media que se corresponden con diseños realizados con uno y dos grupos (tanto si se trata de dos grupos independientes como de dos grupos relacionados). Sin embargo, no siempre nos encontramos con diseños tan simples. Con mucha frecuencia, lo que tenemos que comparar son más de dos grupos pues ello nos dará una idea más amplia de la relación que se pueda establecer entre nuestras variables. En este caso se utiliza el Análisis de Varianza.

El Análisis de Varianza (*Analysis of Variance* o ANOVA) es una técnica de análisis estadístico que se utiliza para comparar las medias de más de dos grupos, aunque su nombre hace referencia al estudio de la variabilidad observada en los datos como veremos a lo largo del tema.

Realmente también existe la prueba *t* de Student, con la que podríamos utilizar pero ello conllevaría una serie de inconvenientes que debemos tener en cuenta. En primer lugar y aunque no es el inconveniente más importante, el número de comparaciones dos a dos que hay que realizar aumenta con el número de grupos o niveles (con tres grupos son tres comparaciones, con cuatro son seis, con cinco son diez, etc.). La regla general para el cálculo del número de comparaciones a realizar son las combinaciones1 del número de grupos tomados de dos en dos (siendo k el número de muestras o grupos, las comparaciones posibles son 2 : $k(k-1))/2$

En segundo lugar, con el aumento del número de comparaciones también aumenta la probabilidad de cometer el error de tipo I (α), esto es, la probabilidad de rechazar la H0 siendo cierta. Podemos considerar este segundo inconveniente como el más grave debido a que toda la Estadística Inferencial clásica se basa en el control del error tipo I. Si nosotros estamos interesados en trabajar con un nivel de riesgo de $\alpha = 0,01$ sabemos que ésta es la probabilidad de rechazar la H0 cierta en cada uno de los contrastes que realicemos y no deseamos que esa probabilidad. se vea aumentada. Pero esto es lo que sucederá si realizamos comparaciones dos a dos. Así, por ejemplo, si realizamos tres contrastes, y asumiendo que los contraste son independientes, la probabilidad de cometer al menos un error de tipo I en los tres vendría dado por

1- (1- α)k = 1- (1- 0,01)3 = 0,0297

Podemos observar que α ha aumentado de 0,01 a 0,03 aproximadamente. Si en lugar de 3 tuviésemos 4 grupos (k=4), la probabilidad de cometer al menos un error de tipo I sería (trabajando con un $\alpha = 0,01$) de 0,0394. Si tuviésemos 5 grupos o muestras (k=5) la probabilidad sería de 0,049. Como vemos, a medida que aumenta el número de grupos a comparar nos estamos alejando del nivel de riesgo inicial con el que queríamos trabajar.

Para superar estos inconvenientes disponemos de una herramienta que realiza la comparación, a la vez, de todos los grupos: el Análisis de Varianza (también conocido como ANOVA y, en algunas publicaciones, como ANVAR), al que le dedicaremos este tema y los dos siguientes.

En este tema vamos a ocuparnos de una de las técnicas de análisis estadístico más utilizadas en Psicología: el Análisis de Varianza (ANOVA o ANVAR). Se trata de introducir al alumno en los fundamentos y la lógica del mismo así como en la utilización de su terminología. Para ello, comenzaremos con el modelo de un solo factor (en temas posteriores se verá el modelo de dos factores), con el que resulta más fácil entender, intuitivamente, la lógica de este tipo de contraste estadístico. Veremos, también, las condiciones que deben cumplir los datos para que se pueda utilizar el Análisis de Varianza e incluiremos, en este tema, el concepto de comparaciones múltiples y una de las técnicas, en este sentido, más utilizadas.

CONCEPTOS BÁSICOS DEL ANÁLISIS DE VARIANZA

Antes de entrar en el fundamento del ANOVA, veamos algunas ideas y terminología que deben ser manejadas con precisión. Siguiendo con nuestro ejemplo, si nosotros comparamos los grupos que han tomado las distintas dosis del fármaco (0, 0,05 mg y 0,10 mg), la variabilidad que aparezca entre ellos puede deberse tanto a los efectos del fármaco como a la influencia de otros factores que no hayamos podido controlar, dado que por muy perfecto que sea el diseño, las variables que existen entre los sujetos y su entorno son tantas que es imposible controlarlas todas. Aunque, por ejemplo, aislásemos e incomunicásemos a los sujetos, esto podría influir de forma distinta en cada uno de ellos. Así pues, a la hora de realizar el estudio hay que ser conscientes de ello, por lo que podemos considerar la variabilidad que se observa entre las puntuaciones, después de haber introducido la variable independiente, como formada por dos partes o componentes:

A) la que se debe al efecto del factor estudiado, en nuestro ejemplo, a las distintas dosis del fármaco.

B) La que se debe a los factores extraños y no controlados, que es lo que recibe el nombre de error experimental, dado que introduce una fuente de error en nuestro diseño. La tarea estará en discernir qué variabilidad corresponde a cada parte y éste es el cometido del Análisis de Varianza.

En la terminología del ANOVA, las variables independientes que se estudian reciben el nombre de factores y las categorías en que se dividen, el de niveles. Así, en nuestro ejemplo, tenemos una sola variable independiente (un solo factor, el fármaco) con tres niveles (0, 0,05 y 0,10 mg, es decir, las distintas dosis suministradas).

Hemos hablado, siguiendo con el ejemplo, de distintas categorías de la variable independiente o niveles: un grupo ha tomado 0,10 mg.; otro 0,05 mg; y el tercero (grupo control) no ha tomado nada (0 mg). Ante estos mismos niveles del factor o variable independiente, nosotros podríamos plantearnos de dos formas la investigación:

a) Nuestro interés puede ser probar que a mayor cantidad del fármaco la ansiedad disminuye.

b) O bien nuestro interés puede ser probar que con 0,10 mg. los sujetos presentan menos ansiedad que con 0,05 mg. y, con esta dosis, menos que sin nada.

Estos dos planteamientos podrían parecer iguales pero no lo son y suponen dos formas distintas de probar la hipótesis nula en el ANOVA. Veamos por qué

Para el primer planteamiento, no nos ceñimos a las dosis concretas del fármaco en cuestión y los niveles actúan como una muestra de todos los posibles niveles que se pudiesen establecer (las posibles dosis que pudiésemos suministrar). A la hora de plantear las conclusiones nos daría lo mismo utilizar estas dosis en el experimento que otras siempre que se pudiese establecer un sentido ascendente o descendente. A este tipo de diseños en los que los niveles son una muestra de los posibles niveles del factor y nuestras conclusiones van a ser para todos ellos, se conoce como diseño de efectos aleatorios o modelo aleatorio.

Para el segundo planteamiento, sin embargo, nuestras conclusiones estarán restringidas a los niveles establecidos en el diseño. Sabemos que pueden existir más niveles pero sólo nos interesan éstos. Este tipo de diseños se conocen como de efectos fijos o modelos fijos.

La interpretación de uno y otro, como vemos, está en función de la hipótesis que se plantea. Más adelante veremos las implicaciones que suponen a la hora del cálculo. En cuanto al número de sujetos, si los grupos o muestras son de distinto número de elementos, el modelo recibe el nombre de modelo no equilibrado y, en el caso de que todas las muestras sean iguales, tendremos un modelo equilibrado.

A- FUNDAMENTOS DEL ANÁLISIS DE VARIANZA

Como hemos visto en la introducción, en ocasiones es necesario comparar más de dos muestras y que, aunque esta comparación podría realizarse de dos en dos por los procedimientos vistos en los temas anteriores (Z o T), el procedimiento resultaría tedioso, no tendría en cuenta la posible interacción entres ellos (el concepto de interacción entre variables independientes o factores se explicará en el siguiente tema) y además aumentaría la probabilidad de cometer el error de tipo I. La técnica a utilizar, en estos casos, puede ser el Análisis de Varianza cuya finalidad es contrastar la diferencia de medias entre varias muestras (dos o más).

La hipótesis que se establece es una hipótesis nula general que consiste en afirmar que no existe diferencia alguna entre las medias (en la variable dependiente) de los distintos grupos o muestras. Sin embargo, como decíamos antes, el modo de contrastar dicha hipótesis va a ser a través de la variabilidad observada en las puntuaciones, de ahí el nombre de esta técnica.

Cuando se diseña un experimento se hace de manera que se intenta minimizar la influencia de las variables extrañas, tanto las conocidas como las desconocidas (todas aquellas variables que están influyendo en la variable dependiente pero que no son la que yo estoy manipulando, que es la única que me interesa). Si no se realiza un control de éstas, se producen sesgos sistemáticos en los resultados que se confunden con los efectos que pudieran deberse a la variable independiente. Un modo de controlar el efecto de variables extrañas desconocidas es la aleatorización de los sujetos, tanto en su obtención (muestreo), como en la asignación a las condiciones experimentales o niveles. Sin embargo, esta aleatorización no asegura que las diferencias observadas entre los niveles no sean fruto del efecto conjunto de la variable independiente y de factores de azar introducidos en el propio proceso. Nunca hay certeza absoluta para atribuir en exclusiva las diferencias entre los niveles o tratamientos sólo al efecto de éstos, y siempre cabe que se atribuya una porción al azar. Algunas de las variables que pueden perturbar y contribuir al error experimental se aleatorizan, pero hay otras muy importantes que no se pueden tratar de la misma manera: diferencias individuales; ambiente en que se desarrolla el experimento; la mejor o peor dicción que tenga el experimentador cuando explica al sujeto lo que hay que hacer, y por tanto que no todos los sujetos lo entiendan de la misma manera; el error de medida; un juicio equivocado sobre la conducta objeto de estudio; error al transcribir los datos del experimento; etc.

Suponemos que todos estos factores influyen de manera no sistemática en el error experimental y que son independientes de los efectos de los tratamientos que, repetimos, son los únicos que nos interesan.

Como se ha señalado, una fuente del error (en el sentido de la existencia de variables extrañas o diferentes a la variable independiente que estoy manipulando y que están afectando a la variable dependiente) son las características (de personalidad, biológicas, etc.) de los propios sujetos. Si tomamos las puntuaciones de todos los sujetos que pertenecen a un mismo grupo o nivel experimental, no es lógico esperar que todas las puntuaciones sean iguales, pues ellos mismos se ven afectados por todas las fuentes de variabilidad no controlada. Por lo tanto, la propia variabilidad de los sujetos sometidos al mismo tratamiento nos proporciona una buena estimación del error experimental. Si extendemos el alcance de este argumento a los sujetos de los demás tratamientos, vemos que se refuerza esta idea de error experimental. Como en principio no podemos suponer que el error sea mayor en un tratamiento que en otro, se puede obtener una mejor y más estable estimación del error (variabilidad entre las puntuaciones) promediando las estimaciones del error que se obtienen para cada tratamiento.

Ahora cambiemos el enfoque y pensemos en los tratamientos por separado, en los cuales las asignaciones de sujetos se han hecho de manera aleatoria. Si la hipótesis nula fuera verdadera (que todas las medias de los tratamientos son iguales), tampoco se nos ocurriría pensar que las medias de los grupos que usamos en el diseño (que son, conviene recordarlo, muestras aleatorias de poblaciones) son necesariamente iguales. Al no ser iguales, y ser verdadera la hipótesis nula de igualdad, de nuevo hay que recurrir a explicar estas diferencias entre las medias de los grupos como un efecto del error experimental. Es decir, todas las fuentes de variabilidad no sistemática que provocan las diferencias entre los sujetos, también influyen en las diferencias que se producen entre los medias muestrales de los grupos.

El Análisis de Varianza, como técnica, le da al investigador "argumentos estadísticos" para decidir si las diferencias que se observan entre los tratamientos son enteramente, o solo parcialmente, debidas al azar. Decíamos en el apartado anterior que la variabilidad observada entre las puntuaciones, en la variable dependiente, debíamos considerarla compuesta por dos componentes:

a) la variabilidad atribuible a los distintos tratamientos experimentales y

b) la variabilidad debida al error experimental.

Pues bien, en el estudio de estas variabilidades se fundamenta el Análisis de Varianza.

En el modelo más simple, el de un sólo factor, se trataría de realizar dos estimaciones, independientes, de la varianza general o común (desconocida) por medio de:

- la varianza atribuible a los distintos niveles del factor en estudio, que es lo que conocemos como varianza intergrupos,

- y por medio de la varianza atribuible al error, que es lo que conocemos como varianza del error o, también, varianza intragrupos.

De la comparación de ambas varianzas obtenemos la aceptación o rechazo de la hipótesis nula.

Tomemos, de nuevo, el ejemplo anterior. Hemos distribuido, aleatoriamente a 15 sujetos en tres grupos, de cinco sujetos cada uno, y les hemos administrado tres dosis distintas de un determinado fármaco para reducir la ansiedad (se supone que los quince sujetos puntuaban alto en esta variable). Dado que los sujetos han sido distribuidos aleatoriamente en los distintos niveles suponemos que, inicialmente, son semejantes en cuanto a la variable estudiada, en todos los niveles. Por lo tanto, si después de suministrarles el tratamiento presentan diferencias hay que pensar que el tratamiento ha influido. Pero, ¿cómo descartar el efecto de factores extraños si, como hemos dicho, la variabilidad que se presenta entre las puntuaciones contiene un componente de error?

El razonamiento del Análisis de Varianza es el siguiente: lógicamente, dentro de un mismo nivel, como el efecto del factor es único (todos los sujetos han tomado la misma dosis del fármaco, en nuestro ejemplo) las puntuaciones deberían ser semejantes, es decir, presentar una variabilidad prácticamente nula. En tanto en cuanto esta variabilidad, dentro de los grupos, sea grande es que están influyendo factores extraños y no controlados en el diseño. Así pues, la varianza dentro de cada grupo o nivel (varianza intragrupos) podemos considerarla como la varianza de error (la que nos mide el error experimental). Por otro lado, si los distintos niveles del factor objeto de estudio producen distintos efectos sobre la variable dependiente, la variabilidad entre los grupos o niveles debería aumentar (la varianza intergrupos). El punto esencial es que si la varianza intergrupos (debida al factor manipulado) es significativamente mayor que la varianza intragrupos (debida a factores de error) se admite que hay diferencias entre los grupos, es decir, que los distintos niveles del factor objeto de estudio han influido de forma distinta sobre la variable dependiente. En nuestro ejemplo, que las distintas dosis del fármaco influyen de forma distinta sobre la ansiedad. La herramienta que tenemos para probar si ambas varianzas (la intergrupos y la intragrupos) difieren significativamente es la distribución F en la que se basa el Análisis de Varianza.

B- ANÁLISIS DE VARIANZA DE UN FACTOR

B-1. Modelo de Efectos fijos.

En este modelo nos interesa estudiar la influencia de un solo factor (al que llamaremos genéricamente factor A o simplemente A), que tiene distintos niveles, como en el ejemplo que estamos comentando: tenemos una variable dependiente, la puntuación en ansiedad en algún test válido y fiable, sobre la que vamos a haceractuar (comprobar el efecto que produce) un factor o variable independiente, que puede presentarse bajo un cierto número de niveles (en este caso las dosis), y nos interesa estudiar, concretamente, el efecto de esos niveles y no otros.

Factor A

	Nivel a_1	Nivel a_2	Nivel a_i	Nivel a_I	Total
	$Y_{1,1}$	$Y_{2,1}$	$Y_{i,1}$	$Y_{I,1}$	
	$Y_{1,2}$	$Y_{2,2}$	$Y_{i,2}$	$Y_{I,2}$	
	$Y_{1,3}$	$Y_{2,3}$	$Y_{i,3}$	$Y_{I,3}$	
	$Y_{1,4}$	$Y_{2,4}$	$Y_{i,4}$	$Y_{I,4}$	
	$Y_{1,j}$	$Y_{2,j}$	$Y_{i,j}$	$Y_{I,j}$	
	Y_{1,n_1}	Y_{2,n_1}	Y_{i,n_1}	Y_{I,n_1}	
Suma	A_1	A_2	A_i	A_I	$T = \sum A$
Media	$\bar{Y}_{A_1}$	$\bar{Y}_{A_2}$	$\bar{Y}_{A_i}$	$\bar{Y}_{A_I}$	$\bar{Y}_T = T/N$
Nº obs.	n_1	n_2	n_i	n_I	$N = \sum n_i$

Consideremos que tenemos I niveles (los niveles del factor en cuestión varían desde $i = 1, \ldots, I$) y n elementos, o sujetos, medidos en cada nivel (los elementos varían desde $j = 1, \ldots, n$ dentro de cada grupo asumiendo un diseño equilibrado). Pues bien, la puntuación Yij que es la puntuación del sujeto que ocupa el lugar j dentro del nivel i, está formada por:

- Un componente al que llamamos αi, y que será común a todos los elementos sometidos a ese nivel del factor.

- El componente del error experimental, formado, según hemos visto, por todos los factores no controlados en el experimento, y que llamamos εij (que puede afectar a todos los sujetos y en todos los niveles, de ahí los dos subíndices).

- Una constante, común para todos los valores de la variable dependiente, que llamaremos μ dado que es la media de la población.

Respecto a los supuestos, de este contraste de hipótesis, hablaremos de ellos a continuación, de modo general, en el apartado 5.7 aunque adelantamos aquí que son tres:

— Independencia de las observaciones.

— Normalidad de las distribuciones.

— Homogeneidad de las varianzas (homocedasticidad).

Estadístico de contraste

El estadístico de contraste para poner a prueba la hipótesis nula de igualdad de medias entre todos los niveles del factor manipulado consiste en elcociente entre la varianza entre los grupos (estimada mediante la media cuadrática inter-grupos o MCinter) y la varianza dentro de los grupos (estimada mediante la media cuadrática intra-grupos o MCintra). Luego veremos cómo se calculan ambas medias cuadráticas. Lo más importante es que, como se estudió en Fundamentos de Investigación, cualquier cociente entre varianzas bajo el supuesto de H0, se distribuirá según la distribución F:

$$F = \frac{MC_{inter}}{MC_{intra}}$$

Es decir, F se distribuye según la distribución *F* de Snedecor con (I - 1) y (N - I) grados de libertad, que son los grados de libertad del numerador y denominador, respectivamente de la razón F en la ecuación 5.4. Veamos cómo llegar a este estadístico de contraste

Vamos a observar la igualdad a la que hemos llegado: la variabilidad total se descompone en la suma de dos variabilidades independientes:

SCtotal = SCInter. + SCintra

SCtotal es la suma de cada puntuación menos la media total al cuadrado.

SCInter. es la suma de la media de cada nivel menos la media total al cuadrado (ponderado por el número de sujetos de cada nivel), es decir, nos está midiendo la variabilidad entre los niveles, o variabilidad debida al efecto del factor (A) que estamos estudiando.

SCintra es la suma de cada puntuación menos la media de su nivel al cuadrado, es decir, nos está midiendo la variabilidad dentro de cada nivel o variabilidad debida al error experimental. Representa las desviaciones de las puntuaciones de cada sujeto respecto de la media de su grupo. Si estas sumas de cuadrados (numeradores de la fórmula de la varianza) las dividimos por sus respectivos grados de libertad, obtendremos: la varianza entre los niveles y la varianza dentro de los niveles.

Los grados de libertad (gl) asociados a las SC se refieren al número de observaciones independientes menos las estimaciones que haya habido que realizar. Los grados de libertad de SCInter. son (I-1) pues tenemos I puntuaciones independientes (los grupos) a los que hay que restar 1 gl por la estimación que realizamos de la media poblacional a partir de la media total. Y los grados de libertad de SCintra son (N-I), pues tenemos N puntuaciones independientes a las que hay que restar las I estimaciones que realizamos de las medias de los grupos.

Por lo que la varianza entre los grupos, a la que llamaremos Media Cuadrática intergrupos o MCinter viene dada por: MCInter. = SCInter./ (I-1)

Tabla 5.2
Tabla del Análisis de Varianza

Fuentes de Variación	**Suma de Cuadrados**	**g.l.**	**Medias Cuadráticas**	**F**
Entre niveles	SC_{inter}	I-1	$MC_{inter} = \frac{SC_{inter}}{I-1}$	$F = \frac{MC_{inter}}{MC_{intra}}$
Dentro de niveles	SC_{intra}	N-I	$MC_{intra} = \frac{SC_{intra}}{N-I}$	
Total	SC_{Total}	N-1		

Ejemplo 5.3. Veamos un ejemplo: supongamos que tenemos una muestra aleatoria de nueve sujetos que puntúan alto en ansiedad y los distribuimos aleatoriamente en tres grupos de forma que $n_1 = 4$, $n_2 = 3$ y $n_3 = 2$. Les suministramos tres dosis distintas de una cierta droga: 0,05 mg; 0,10 mg; y 0,20 mg. Queremos ver si estas dosis influyen en el estado de ansiedad de los sujetos. Se cumplen los supuestos de independencia de las observaciones, normalidad de las distribuciones y homocedasticidad. Los resultados fueron:

$a_1 = 0,05$	$a_2 = 0,10$	$a_3 = 0,20$
5	6	2
8	7	4
4	8	
3		

Pasos del contraste:

1.- En primer lugar, ¿cuáles son las hipótesis estadísticas que queremos probar?

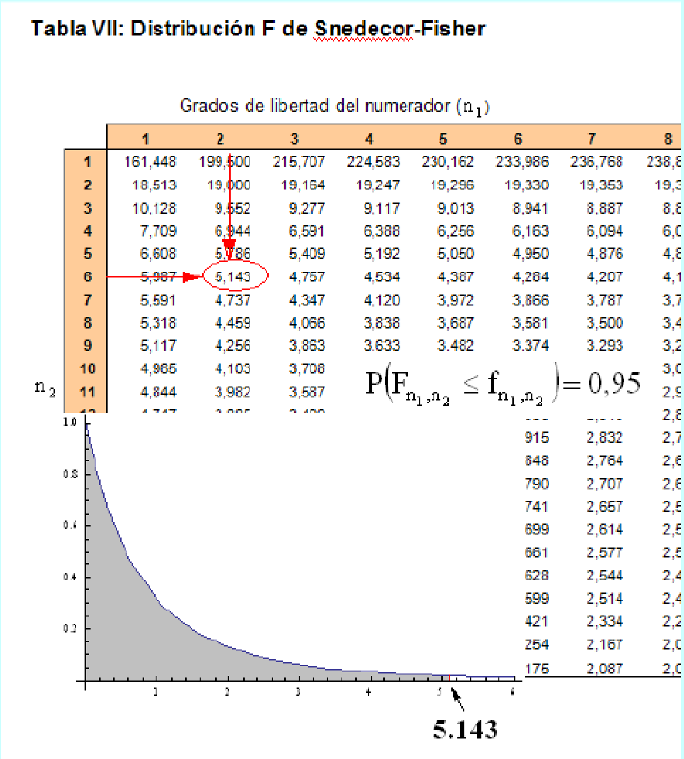
H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: $\mu_1 \neq \mu_2 \neq \mu_3$ al menos para una μ_i

2.- En segundo lugar, deberemos probar los supuestos de independencia de las observaciones, normalidad de las distribuciones y homogeneidad de varianzas, para comprobar que puede aplicarse el Análisis de Varianza. En este caso el enunciado nos dice que se cumplen.

3.- El tercer paso será decidir cuál es el estadístico de contraste que vamos a utilizar para probar la hipótesis nula.

$$F = \frac{MC_{inter}}{MC_{intra}}$$

4.- El cuarto paso será decidir con qué nivel de confianza queremos trabajar y, en función de ese nivel, establecer la región crítica de rechazo de la hipótesis nula. En este caso, si trabajamos con un nivel de confianza del 95%, el valor crítico de las tablas de la distribución F de Snedecor para (I-1) = (3-1) = 2 grados de libertad en el numerador y (N-I) = (9-3) = 6 grados de libertad en el denominador, es 5,143.



n_2	1	2	3	4	5	6	7	8
1	161,448	199,500	215,707	224,583	230,162	233,986	236,768	238,8
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,3
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,8
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,0
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,8
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,1
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,7
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,4
9	5,117	4,256	3,863	3.633	3.482	3.374	3.293	3,2
10	4,965	4,103	3,708					3,0
11	4,844	3,982	3,587					2,9

5.- A continuación calculamos el estadístico:

Niveles	a_1	a_2	a_3	
Dosis	0,05	0,10	0,20	
	$Y_{11}= 5$	$Y_{21}= 6$	$Y_{31}= 2$	
	$Y_{12}= 8$	$Y_{22}= 7$	$Y_{32}= 4$	
	$Y_{13}= 4$	$Y_{23}= 8$		
	$Y_{14}= 3$			
n_i	4	3	2	N = 9

Σ	$A_1 = 20$	$A_2 = 21$	$A_3 = 6$	$\Sigma A = T = 47$
$\bar{Y}_{A_i}$	5	7	3	$\bar{Y}_T = 5{,}22$

$$SC_{inter} = \sum_{i=1}^{I} n_i(\bar{Y}_{A_i} - \bar{Y}_T)^2 = 4(5\text{-}5{,}22)^2 + 3(7\text{-}5{,}22)^2 + 2(3\text{-}5{,}22)^2 = 19{,}5556$$

$$SC_{intra} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{A_i})^2 = (5\text{-}5)^2 + (8\text{-}5)^2 + (4\text{-}5)^2 + (3\text{-}5)^2 + (6\text{-}7)^2 + (7\text{-}7)^2 + (8\text{-}7)^2 + (2\text{-}3)^2 + (4\text{-}3)^2 = 18$$

$$SC_{Total} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_T)^2 = (5\text{-}5{,}22)^2 + (8\text{-}5{,}22)^2 + (4\text{-}5{,}22)^2 + (3\text{-}5{,}22)^2 + (6\text{-}5{,}22)^2 + (7\text{-}5{,}22)^2 + (8\text{-}5{,}22)^2 + (2\text{-}5{,}22)^2 + (4\text{-}5{,}22)^2 = 37{,}5556$$

Como podemos comprobar, $SC_{Total} = SC_{inter} + SC_{intra}$;
$37{,}5556 = 19{,}5556 + 18$

Tabla del Análisis de Varianza:

Fuentes de Variación	**Suma de Cuadrados**	**g.l.**	**Medias Cuadráticas**	**F**
Entre niveles	19,5556	3-1	19,5556 / 2 = 9,778	$F = \boxed{\frac{9{,}778}{3} = 3{,}259}$
Dentro de niveles	18	9-3	18 / 6 = 3	
Total	37,5556	9-1		

6.- Una vez calculado el estadístico de contraste F, el siguiente paso será tomar la decisión sobre la hipótesis nula, comparando el valor obtenido con el valor crítico que establecíamos en el 4º paso y observamos que el estadístico de contraste no supera el valor crítico:

$$3{,}259 < 5{,}14$$

por lo tanto, no podemos rechazar la hipótesis nula.

7.- El último paso del análisis será la interpretación: dado que no hemos podido rechazar la hipótesis nula nuestra interpretación es que, a un nivel de confianza del 95%, no podemos rechazar que las tres medias son iguales y las diferencias que observamos entre ellas se deben al azar, o lo que es lo mismo, que los tratamientos suministrados no parecen influir de forma distinta sobre la ansiedad.

Hemos visto en la resolución del problema anterior cómo si el número de sujetos, o de grupos fuese un poco mayor, los cálculos a realizar serían considerablemente

Vamos a aplicar este método abreviado a los mismos datos del ejemplo anterior:

Calculamos el estadístico:

	Niveles						
	a_1		a_2		a_3		
Dosis	0,05		0,10		0,20		
	Puntua.	Cuadrados	Puntua.	Cuadrados	Puntua.	Cuadrados	
	Y_{11}= 5	25	Y_{21}= 6	36	Y_{31}= 2	4	
	Y_{12}= 8	64	Y_{22}= 7	49	Y_{32}= 4	16	
	Y_{13}= 4	16	Y_{23}= 8	64			
	Y_{14}= 3	9					
Σ	A_1 = 20	114	A_2 = 21	149	A_3 = 6	20	ΣA = T = 47
n_i	4		3		2		N = 9
$\bar{Y}_A$	5		7		3		$\bar{Y}_T$ = 5,22

$$SC_{inter} = \sum_i \frac{(A_i)^2}{n_i} - \frac{T^2}{N}$$

$$= \frac{20^2}{4} + \frac{21^2}{3} + \frac{6^2}{2} - \frac{47^2}{9} = 19{,}5556$$

$$SC_{Total} = \Sigma\Sigma\, Y^2 - \frac{T^2}{N} =$$