2011

UNED

DISEÑOS DE INVESTIGACIÓN Y ANÁLISIS DE DATOS

[TEMA 5]

Diseños con más de dos grupos independientes

Índice

5.1.- INTRODUCCIÓN

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 (***An****alysis* ***of*** ***Va****riance* 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.

Antes de introducirnos en esta técnica de análisis veamos un ejemplo que nos ayude a entender la lógica y los mecanismos de este procedimiento. En el Tema 3 (ejemplo 3.3, apartado 3.4.4) veíamos un ejemplo en el que queríamos probar el efecto del consumo de un determinado fármaco en la reducción de la ansiedad y teníamos un grupo experimental y un grupo control. A la hora de analizar los resultados, realizamos un contraste de medias y, en este caso, el estadístico de contraste (t = - 2,46) sigue una distribución *t-Student* con 8 grados de libertad ($n_1 + n_2 - 2$).

Dado que la probabilidad de encontrar un valor igual o menor al que hemos obtenido es inferior a 0,05 decidimos rechazar la hipótesis nula de igualdad de medias poblacionales, concluyendo que el fármaco en cuestión sí influye en la ansiedad que presentan los sujetos.

Ahora bien, en lugar de limitarnos a este experimento tan sencillo podríamos ir un poco más allá y estar interesados, no sólo en la existencia de diferencias en los grupos como consecuencia de la administración o no del fármaco, sino que además nos puede interesar averiguar si existe un efecto distinto en función, por ejemplo, de dos cantidades distintas del fármaco suministrado.

De este modo, podemos complicar un poco más el diseño y, en lugar de dos grupos, podemos construir aleatoriamente tres grupos: el primero, sin tomar el fármaco; el segundo, con una dosis de 0,05 miligramos; y el tercero, con una dosis de 0,10 miligramos. Supongamos que en esta ocasión los datos que hemos obtenido son los del ejemplo 5.1 que aparece a continuación.

Ejemplo 5.1. En este caso vamos a trabajar con tres grupos porque, además de estar interesados en si el fármaco influye o no en la ansiedad, queremos averiguar si influye de forma distinta en función de que se le suministre 0,05 o 0,10 mg.

Supongamos que hemos obtenido los siguientes resultados:

	Sin fármaco	**Con 0,05 mg de fármaco**	**Con 0,10 mg de fármaco**
	30	10	20
	50	20	40
	30	30	10
	60	20	10
	20	5	10
n_j	5	5	5
$\sum$	190	85	90
$\bar{Y}_j$	38	17	18
$\hat{S}_j^2$	270	95	170

Nota: recuérdese que el acento circunflejo sobre el símbolo de la varianza muestral representa la varianza insesgada muestral o cuasivarianza.

Disponemos de una herramienta conocida, la prueba *t* de Student, 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 **combinaciones**[1] del número de grupos tomados de dos en dos (siendo k el número de muestras o grupos, las comparaciones posibles son $k(k-1)/2$). En nuestro caso, como tenemos 3 muestras, el número de comparaciones serían tres, es decir $3(3-1)/2 = 3$.

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 H_0 siendo cierta. Podemos considerar este segundo inconveniente como el más grave debido a que toda la Estadística Inferencial clásica[2] 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 H_0 cierta en cada uno de los contrastes que realicemos y no deseamos que esa probabilidad

[1] Dado un conjunto A de m elementos, se denomina combinaciones de orden n (n<m) a las distintas agrupaciones de n elementos elegidos del conjunto A que pueden realizarse de modo que dos agrupaciones son distintas si se distinguen únicamente por uno o más de los elementos que la componen. Es decir, dos subconjuntos de n elementos se consideran la misma combinación si sólo difieren en el orden pero no en la identidad de los elementos.

[2] Existen otras alternativas a la inferencia estadística "clásica", la más importante de las cuales es la Inferencia Bayesiana. Este curso no puede entrar en su explicación.

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

$$1-(1-\alpha)^k = 1-(1-0{,}01)^3 = 0{,}0297$$

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 α = 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.

5.2.- OBJETIVOS DEL TEMA

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.

5.3. 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**.

5.4. 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.

5.5 ANÁLISIS DE VARIANZA DE UN FACTOR

5.5.1. 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 hacer actuar (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.

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 Y_{ij} que es la puntuación del sujeto que ocupa el lugar j dentro del nivel i, está formada por:

- Un componente al que llamamos α_i, 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 ε_{ij} (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.

La acción de estas tres componentes se supone lineal, de forma que el modelo sería:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad 5.1$$

Donde:

ε_{ij} (el componente de error) es una variable aleatoria distribuida según $N(0, \sigma)$

$\sum_{i=1}^{I} n_i \alpha_i = 0$ bajo la H_0 de que no hay efecto.

Esta es, pues, la estructura de una observación cuando estamos analizando un solo factor.

Antes de continuar con la exposición, veamos la **terminología** estadística que vamos a utilizar:

TABLA 5.1
Sistema de notación para el ANOVA de un factor

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	$N = \sum n_i$

- Hemos visto que Y_{ij} es la puntuación (en la variable dependiente) que obtiene el sujeto *j* en el nivel *i*.
- Que α_i es un elemento común a todos los sujetos experimentales sometidos a un mismo nivel del factor, y es el efecto que queremos medir.
- Que n_i es el número de sujetos u observaciones en el nivel *i*.
- N es el número total de sujetos.
- $\bar{Y}_{A_i}$ es la media aritmética de cada nivel (*i*). Es el resultado de sumar todas las puntuaciones, de cada nivel y dividirlo por el número de sujetos de ese nivel:

$$\bar{Y}_{A_i} = 1/n_i \sum_{j=1}^{n_i} Y_{ij}$$

- $\overline{Y}_T$ es la media aritmética total, o media de todas las puntuaciones. Es decir, el resultado de sumar todas las puntuaciones de cada nivel y en todos los niveles, y dividirlo por el número total de sujetos.

$$\overline{Y}_T = 1/N \sum_{i=1}^{I} \sum_{j=1}^{ni} Y_{ij}$$

Se demuestra que la media total ($\overline{Y}_T$)es una estimación insesgada de la media de la población μ:

$$\overline{Y}_T \approx \mu \qquad 5.2$$

Y que la media de cada nivel es una estimación sesgada de la media de la población, con un sesgo aditivo consistente en el efecto del factor:

$$\overline{Y}_{A_i} \approx \mu + \alpha_i \qquad 5.3$$

Por lo tanto, si restamos la expresión 5.2 de la expresión 5.3:

$$\overline{Y}_{A_i} - \overline{Y}_T \approx \mu + \alpha_i - \mu$$

Luego

$$\overline{Y}_{A_i} - \overline{Y}_T \approx \alpha_i$$

¿Cuándo diremos que los tratamientos, o niveles del factor, influyen de forma distinta sobre la variable dependiente? Lo que nos estamos preguntando es si los α_i son iguales dado que, (recordemos) α_i es el elemento común a todos los sujetos experimentales sometidos a un mismo nivel del factor que estemos estudiando. Es decir, queremos saber si:

$\alpha_1 = \alpha_2 = \alpha_3 = \ldots. = \alpha_I$

Pero, como hemos dicho que $\sum_{i=1}^{I} n_i\alpha_i = 0$ bajo la H_0 de que no hay efecto, lo fundamental es probar si alguno de los α_i es igual a cero:

$$\alpha_1 = \alpha_2 = \alpha_3 = \ldots. = \alpha_I = 0$$

dado que, en ese caso, el efecto del tratamiento será nulo.

Las **hipótesis estadísticas** que queremos probar en este tipo de problemas son:

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

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 el cociente entre la varianza entre los grupos (estimada mediante la media cuadrática inter-grupos o MC_{inter}) y la varianza dentro de los grupos (estimada mediante la media cuadrática intra-grupos o MC_{intra}). 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 H_0, se distribuirá según la distribución F:

$$\boldsymbol{F = \frac{MC_{inter}}{MC_{intra}}} \qquad \textbf{5.4}$$

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.

Variabilidad del sistema
Realizamos el doble sumatorio, para todos los niveles (uno por uno) y para todos los sujetos que hay en cada nivel, de cada puntuación menos la media total, elevada al cuadrado, es decir de los N sujetos. Si recordamos la fórmula de la varianza, esto sería el numerador de la varianza total.

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_T)^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_T+\overline{Y}_{A_i}-\overline{Y}_{A_i})^2$$

En el segundo término de la igualdad no hemos hecho más que sumar y restar la media de cada grupo (con lo que no se altera el sumatorio).

Pues bien, reagrupando el segundo término de la igualdad, tenemos:

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_T+\overline{Y}_{Ai}-\overline{Y}_{A_i})^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}\left[(Y_{ij}-\overline{Y}_{A_i})+(\overline{Y}_{A_i}-\overline{Y}_T)\right]^2 =$$

$$= \sum_{i=1}^{I} n_i(\overline{Y}_{A_i}-\overline{Y}_T)^2 + \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{A_i})^2$$

dado que se trata del cuadrado de una suma y se demuestra que el doble producto es igual a cero, con lo que:

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_T)^2=\sum_{i=1}^{I}n_i(\bar{Y}_{A_i}-\bar{Y}_T)^2+\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{A_i})^2$$

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

$$\boxed{SC_{total} = SC_{Inter.} + SC_{intra}}$$

SC_{total} es la suma de cada puntuación menos la media total al cuadrado.

$SC_{Inter.}$ 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.

SC_{intra} 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.

Recordemos que:

$$\text{Varianza} = \frac{\textit{suma de cuadrados de las desviaciones de la media}}{\textit{grados de libertad}}$$

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 $SC_{Inter.}$ 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 SC_{intra} 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 MC_{inter} viene dada por: $MC_{Inter.} = SC_{Inter.} / (I-1)$

Y la varianza dentro de los grupos, a la que llamaremos Media Cuadrática intragrupos o MC_{intra}, viene dada por: $MC_{intra} = SC_{intra} / (N-I)$.

Decíamos al principio del tema, al hablar de la lógica del Análisis de varianza que en el modelo más simple, el de un sólo factor, se trata 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**.

Y que de la comparación de ambas varianzas obtenemos la aceptación o rechazo de la hipótesis nula. Pues bien, ya vemos cómo se calculan dichas varianzas, nos faltaría especificar cómo compararlas.

Se demuestra que la esperanza matemática de la varianza dentro de los niveles (la media cuadrática intragrupos), es un estimador insesgado de la varianza poblacional, mientras que la varianza entre los niveles (media cuadrática intergrupos) es un estimador sesgado de la varianza poblacional, con un sesgo aditivo (un sumando) función del efecto del factor (al que hemos llamado α_i), es decir, cuanto más efecto ejerza la variable independiente sobre la dependiente, mayor será el sesgo en la estimación de MC_{Inter}:

$$E(MC_{Inter.}) = \sigma^2 + \frac{1}{(I-1)}\sum_{i=1}^{I} n_i\alpha_i^2$$

$$E(MC_{intra}) = \sigma^2$$

Por lo tanto si dividimos la $MC_{Inter.}$ entre la MC_{intra} el cociente deberá ser igual o mayor de uno.

Si MC_{Inter} es muy grande (más de lo que cabría esperar por efecto del azar en el muestreo) respecto a MC_{intra} admito que algún $\alpha_i \neq 0$ pues, en teoría, si todos los $\alpha_i = 0$ las dos esperanzas que componen la razón entre MC_{Inter} y MC_{intra} son iguales y por lo tanto el cociente igual a la unidad.

¿Cómo determinar si MC_{Inter} es o no muy grande respecto a MC_{intra}? A través de la distribución F, pues dicho cociente se demuestra que se distribuye según dicha distribución con los grados de libertad del numerador y del denominador.

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

La razón F, como decíamos unas páginas atrás, es el estadístico de contraste del Análisis de Varianza.

Tabla del Análisis de Varianza.

Los resultados de un Análisis de Varianza suelen presentarse en tablas del tipo que puede verse en la Tabla 5.2:

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?

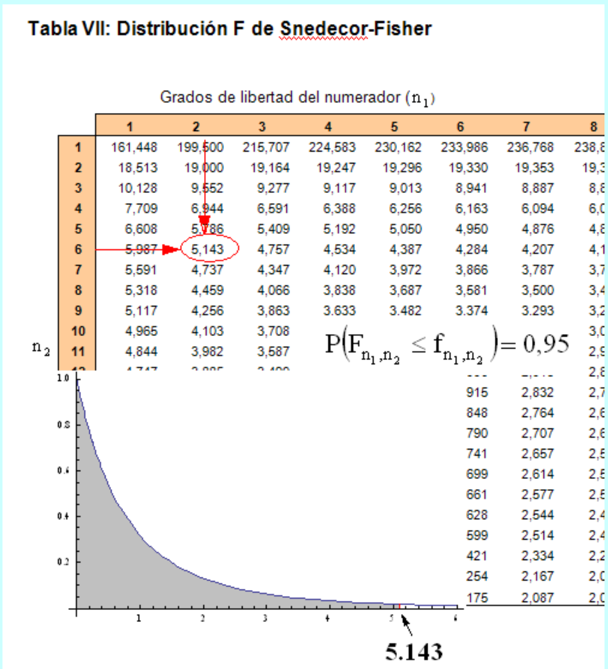
H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: $\mu_1 \neq \mu_2 \neq \mu_3$ al menos para una μ_i

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.



	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	$\bar{Y}_T = 5{,}22$

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

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

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

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 largos. Normalmente se acude a programas informáticos para realizarlos (SPSS, SAS,…) o bien, a partir del desarrollo de las fórmulas de cálculo de SC_{Total} y de $SC_{intra.}$ podemos llegar a unos cálculos abreviados en base a la terminología vista en la tabla 5.1

$$SC_T = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_T)^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}Y_{ij}^{\ 2} - \frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{n_i}Y_{ij}\right)^2}{N} = \Sigma\Sigma\, Y^2 - \frac{T^2}{N}$$

$$SC_{inter} = \sum_{i=1}^{I} n_i(\bar{Y}_{A_i} - \bar{Y}_T)^2 = \left[\sum_{i=1}^{I}\frac{\left(\sum_{j=1}^{n_i}Y_{ij}\right)^2}{n_i}\right] - \frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{n_i}Y_{ij}\right)^2}{N} = \sum_i \frac{(A_i)^2}{n_i} - \frac{T^2}{N}$$

$$SC_{intra} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{A_i})^2 = SC_T - SC_{inter}$$

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

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

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

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

$= (114+149+20) - \frac{47^2}{9} = 37{,}5556$

$SC_{intra} = SC_{Total} - SC_{inter} = 37{,}5556 – 19{,}5556 = 18$

Como podemos comprobar, los resultados son los mismos.

5.5.2. Modelo de Efectos aleatorios.

Todo lo dicho hasta ahora, en el apartado anterior sobre el modelo de un factor, se refería al modelo de efectos fijos. Decíamos en el apartado de conceptos básicos que un modelo es de efectos fijos cuando el investigador establece como niveles del factor sólo aquellos que está interesado en estudiar. Si el modelo es de efectos aleatorios, decíamos que los *i* niveles del factor son una muestra aleatoria de todos los posibles niveles. Esta muestra aleatoria de niveles se supone distribuida $N(0, \sigma_A)$, siendo A el factor en cuestión.

Aunque el planteamiento es distinto, cuando se trata de un solo factor este hecho no tiene consecuencias para el cálculo. El estadístico de contraste sigue siendo el mismo:

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

donde F se distribuye según la distribución F de Snedecor con (I-1) y (N-I) grados de libertad.

5.6. COMPARACIONES MÚLTIPLES.

Un objetivo de toda investigación está en obtener el máximo de resultados e interpretaciones a partir de las observaciones de que disponemos. Cuando un investigador aplica un Análisis de Varianza a un conjunto de observaciones para detectar si hay diferencias entre los distintos grupos sometidos a tratamientos distintos, o lo que es lo mismo, si hay diferencias entre los resultados de la variable dependiente en función de distintos niveles de la variable independiente (o de distintos niveles de distintas variables independientes como veremos en el tema siguiente), lo único que obtiene es un razón F que, en caso de ser un resultado significativo (es decir que no podamos mantener la H_0 de que todas las medias son iguales) nos lleva a una H_1 que indica que **al menos** entre dos medias hay diferencias que no son debidas al azar. Pero, no nos dice dónde está esa diferencia significativa.

Las comparaciones múltiples permiten establecer una información más exacta sobre la importancia de cada uno de los niveles de la variable independiente.

Al plantearnos el análisis de las comparaciones múltiples entre las medias de los distintos niveles de la variable independiente, cabe distinguir dos situaciones básicas:

1.- La primera se refiere a la situación más común en la que el investigador, una vez realizado el Análisis de Varianza y rechazada la H_0 desea conocer entre qué medias hay diferencias no debidas al azar. Se trata de las comparaciones **no planificadas, a posteriori o *post hoc***.

2.- La segunda se refiere a cuando el investigador no está interesado en realizar un Análisis de Varianza para probar todas las medias sino sólo en algunas comparaciones entre los niveles del factor, no en todas las posibles y sabe de antemano qué comparaciones le interesan. Se trata de **comparaciones planificadas o *a priori***.

El objeto de las comparaciones múltiples es, como parte del Análisis de Varianza, reducir la cantidad de error Tipo I que cometeríamos si comparásemos dos a dos todas las muestras, por lo tanto, aunque comparemos las muestras dos a dos, no recurrimos a la prueba *t* estudiada en temas precedentes, sino que aplicaremos pruebas específicas que aprovechan los resultados del Análisis de Varianza y que nos aseguran que no se incrementa el error de tipo I (α).

5.6.1. COMPARACIONES PLANIFICADAS O *A PRIORI*

Como decíamos en el apartado 5.6, en ocasiones el investigador no está interesado en comparar dos a dos todas las medias, sino sólo algunas combinaciones de los niveles del factor pero, dado que si se realizan comparaciones dos a dos con la prueba *t* se incrementa el error de tipo I (según hemos visto al inicio del tema) realiza un Análisis de Varianza, aunque no es su resultado el que le interesa, sino esas comparaciones respecto a las hipótesis específicas diseñadas de antemano. De ahí su nombre de comparaciones planificadas o *a priori*. De hecho, en los resultados de la investigación no suele citarse, siquiera, el valor F obtenido en el análisis de varianza.

Ejemplo 5.4 Supongamos que estamos interesados en conocer el efecto que sobre la reducción de la ansiedad tiene el realizar terapias conductuales (vamos a suponer 4 terapias distintas). No nos interesa comparar las terapias entre sí. Tomamos a un grupo de sujetos, que puntúan semejante en ansiedad, y los distribuimos aleatoriamente en cinco grupos: uno de ellos será el grupo control y los otros cuatro serán tratados cada uno de ellos con una terapia distinta. Al finalizar la terapia medimos la variable dependiente (ansiedad) de los cinco grupos.

Tendríamos que comparar, para probar nuestra hipótesis, las puntuaciones del grupo control con las de los otros cuatro grupos (cuatro comparaciones).

Hemos visto en la Introducción del tema cómo a medida que aumenta el número de grupos a comparar nos alejamos del nivel de riesgo inicial con el que queríamos trabajar (aumentamos la probabilidad de cometer error de tipo I). Es por ello que recurrimos al Análisis de Varianza, aunque desechemos parte de sus resultados.

Dado el alcance del curso, no vamos a detenernos en este tipo de comparaciones, pero sí apuntarlas para que el estudiante conozca su existencia.[3]

5.6.2. COMPARACIONES NO PLANIFICADAS, A POSTERIORI O *POST HOC*

Las comparaciones no planificadas, *a posteriori* o *post hoc* (simplemente comparaciones múltiples para algunos autores) son aquellas, como hemos dicho, que se deciden después de que el investigador haya obtenido los resultados del Análisis de Varianza, rechazando la hipótesis nula. Aunque existen distintas técnicas para realizar estas comparaciones, aquí vamos a estudiar sólo una de ellas: la prueba de comparaciones múltiples de Scheffé, que es una de las más utilizadas.

5.6.2.1. Prueba de comparaciones múltiples de Scheffé.

Esta prueba propuesta por Scheffé permite no sólo comparar las medias de los niveles del Análisis de Varianza dos a dos, sino también realizar comparaciones complejas, esto es: comparar la media de un nivel con un conjunto de medias de otros niveles; o comparar un conjunto de medias de distintos niveles con otro conjunto de medias de otros niveles.

Esta prueba fija la tasa de error de tipo I en el α al que estemos trabajando, sin aumentarlo en todas las posibles comparaciones que realicemos, y obtiene un valor al que llama diferencia mínima o rango crítico (***C****ritical* ***R****ange* de Scheffé) por encima de la cual diremos que hay diferencias entre las medias o entre los grupos de medias que estemos comparando. Esta diferencia mínima se calcula según la fórmula:

[3] Remitimos al estudiante interesado en el tema de las comparaciones múltiples al libro clásico de Keppel (1973) *Design and anlysis: A researcher's handbook*. Editado por Prentice Hall, Inc.

$$\mathrm{CR_{Scheffé}} = \sqrt{(I-1)F_{(i-1),gl_{error}}}\sqrt{MC_{error}\left(\sum_{i=1}^{I}\left(c_i^2 / n_i\right)\right)}$$

Donde c_i son los coeficientes de las combinaciones lineales que podemos establecer entre las distintas medias a comparar, y el resto de valores se puede obtener de la tabla del ANOVA. En cada combinación de coeficientes c_i, la suma de los mismos debe ser igual a cero. Por ejemplo, si tenemos tres niveles en el factor y queremos hacer todas las comparaciones posibles dos a dos de las medias, resultarán tres comparaciones y los coeficientes serían:

	a_1	a_2	a_3	Medias a comparar
	1	-1	0	$\bar{Y}_1 - \bar{Y}_2$
Coeficientes	1	0	-1	$\bar{Y}_1 - \bar{Y}_3$
	0	1	-1	$\bar{Y}_2 - \bar{Y}_3$

Observemos que cuando uno de los coeficientes es 0, esto significa que elimina a ese grupo de la comparación. Cuando las comparaciones implican a más de dos grupos, los valores de los coeficientes deben reflejar los grupos a comparar y el tipo de comparación. En la Tabla 5.3, se muestran algunos ejemplos de comparaciones en un diseño de un factor con 5 niveles.

Tabla 5.3

Ejemplos de Coeficiente para comparación de medias de grupos

	Hipótesis sobre comparaciones	a_1	a_2	a_3	a_4	a_5	$\sum c_i$
		Coeficientes c_i					
Ej. 1	$H_0: \mu_1 = \frac{\mu_2 + \mu_3}{2}$	2	-1	-1	0	0	0
Ej. 2	$H_0: \mu_5 = \frac{\mu_2 + \mu_4}{2}$	0	-1	0	-1	2	0
Ej. 3	$H_0: \frac{\mu_1 + \mu_5}{2} = \frac{\mu_2 + \mu_3 + \mu_4}{3}$	3	-2	-2	-2	3	0
Ej. 4	$H_0: \mu_2 = \frac{\mu_3 + \mu_4 + \mu_5}{3}$	0	3	-1	-1	-1	0
Ej. 5	$H_0: \mu_1 = \frac{\mu_2 + \mu_3 + \mu_4 + \mu_5}{4}$	4	-1	-1	-1	-1	0

Observe el estudiante que los coeficientes están relacionados con el número de grupos en cada lado de la expresión de H_0. Así por ejemplo, en el Ej. 3, el más complicado quizás, se trata de comparar si la media de las medias de los grupos 1 y 5 es igual a la media de las medias de los grupos 2, 3 y 4. La misma expresión formulada en la hipótesis nula, se puede escribir de la siguiente manera:

$$H_0: 3(\mu_1 + \mu_5) - 2(\mu_2 + \mu_3 + \mu_4) = 0$$

Si se multiplican los factores 2 y 3 por la expresión dentro de su correspondiente paréntesis, se ve claramente el valor de los coeficientes que afectan a cada media de grupo.

Ejemplo 5.5 Supongamos que tenemos una muestra aleatoria de 21 sujetos que puntúan alto en ansiedad y los distribuimos aleatoriamente en tres grupos de 7 sujetos cada uno. 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.
Una vez realizado el Análisis de varianza, los resultados fueron:

$\bar{Y}_1 = 11{,}7143 \qquad \bar{Y}_2 = 15{,}4 \qquad \bar{Y}_3 = 8{,}57$

F.V.	**S.C.**	**g.l.**	**M.C.**	**F**
Entre niveles	164,95	3-1	$\frac{164{,}95}{2} = 82{,}475$	$\frac{82{,}475}{16{,}27} = 5{,}069$
Dentro niveles	292,86	21-3	$\frac{292{,}86}{18} = 16{,}27$	
Total	457,81	21-1		

Si acudimos a las tablas de la distribución F, el valor crítico para 2 y 18 grados de libertad, trabajando con un nivel de confianza del 95%, es 3,555.

Tabla VII: Distribución F de Snedecor-Fisher

$$P\left(F_{n_1,n_2} \le f_{n_1,n_2}\right) = 0{,}95$$

n_2	1	2	3	4	5
1	161,448	199,500	215,707	224,583	230,162
2	18,513	19,000	19,164	19,247	19,296
3	10,128	9,552	9,277	9,117	9,013
4	7,709	6,944	6,591	6,388	6,256
5	6,608	5,786	5,409	5,192	5,050
6	5,987	5,143	4,757	4,534	4,387
7	5,591	4,737	4,347	4,120	3,972
8	5,318	4,459	4,066	3,838	3,687
9	5,117	4,256	3,863	3,633	3,482
10	4,965	4,103	3,708	3,478	3,326
11	4,844	3,982	3,587	3,357	3,204
12	4,747	3,885	3,490	3,259	3,106
13	4,667	3,806	3,411	3,179	3,025
14	4,600	3,739	3,344	3,112	2,958
15	4,543	3,682	3,287	3,056	2,901
16	4,494	3,634	3,239	3,007	2,852
17	4,451	3,592	3,197	2,965	2,810
18	4,414	3,555	3,160	2,928	2,773
19	4,381	3,522	3,127	2,895	2,740
20	4,351	3,493	3,098	2,866	2,711
30	4,171	3,316	2,922	2,690	2,534
60	4,001	3,150	2,758	2,525	2,368

Comparando nuestro resultado con el de las tablas vemos que el estadístico de contraste supera al nivel crítico (5,069 > 3,55) por lo que rechazaremos la H_0 de igualdad de medias. Ahora bien, ¿entre qué pares de medias está la diferencia que hace que rechacemos la hipótesis nula?

Si aplicamos la prueba de comparaciones múltiples de **Scheffé:**

$$CR_{Scheffé} = \sqrt{(I-1)F_{(I-1),glerror}}\sqrt{MC_{error}\left(\sum_{i=1}^{I}\left(c_i^2 / n_i\right)\right)}$$

a) $n_1 = n_2 = n_3 = n_i = 7$ $I = 3$ $\bar{Y}_1 = 11{,}7143$ $\bar{Y}_2 = 15{,}4$ $\bar{Y}_3 = 8{,}57$

b) El valor de la distribución F para un $\alpha = 0{,}05$; con 2 grados de libertad en el numerador y 18 grados de libertad en el denominador, es igual a 3,555.

c) $MC_{error} = 16{,}27$

d) Si queremos comparar todas las medias entres sí, tenemos tres posibles comparaciones, y establecemos tres combinaciones lineales[4]:

1ª: Para comparar la media 1 con la media 2: $(1)\bar{Y}_1 + (-1)\bar{Y}_2 + (0)\bar{Y}_3 = (1)\bar{Y}_1 + (-1)\bar{Y}_2$

2ª: Para comparar la media 1 con la media 3: $(1)\bar{Y}_1 + (0)\bar{Y}_2 + (-1)\bar{Y}_3 = (1)\bar{Y}_1 + (-1)\bar{Y}_3$

3ª: Para comparar la media 2 con la media 3: $(0)\bar{Y}_1 + (1)\bar{Y}_2 + (-1)\bar{Y}_3 = (1)\bar{Y}_2 + (-1)\bar{Y}_3$

e) Calculamos $CR_{Scheffé} = \sqrt{(2)(3{,}55)}\sqrt{(16{,}27)((1/7)+(1/7))} = 5{,}745$

Esta es la diferencia mínima o rango crítico (***C****ritical* ***R****ange* de Scheffé), por encima de la cual diremos que hay diferencias entre las medias o entre los grupos de medias que estemos comparando. Por lo tanto, realizamos las comparaciones:

$|\bar{Y}_1 - \bar{Y}_2| = |11{,}7143 - 15{,}43| = 3{,}7157 < 5{,}745$

$|\bar{Y}_1 - \bar{Y}_3| = |11{,}7143 - 8{,}57| = 3{,}1443 < 5{,}745$

$|\bar{Y}_2 - \bar{Y}_3| = |15{,}43 - 8{,}57| = 6{,}86 > 5{,}745$

Como vemos, la única diferencia significativa (responsable de que hayamos rechazado la hipótesis nula del análisis de varianza) se da entre los grupos 2 y 3 ya que la diferencia de medias entre estos dos grupos supera el valor del CR de Scheffé.

Decíamos que una de las posibilidades que ofrece la prueba de Scheffé es que nos permite realizar comparaciones más complejas que la simple comparación, dos a dos, de las muestras. Veamos un ejemplo.

Ejemplo 5.6. Supongamos, siguiendo con el mismo ejemplo, que de las tres dosis distintas la primera (0,05) a la que llamaremos *A*, es la dosis tradicional que se

[4] Recordemos que, en cada combinación, la suma de los coeficientes debe ser igual a cero.

suele suministrar mientras que las otras dos, a las que llamaremos *B* y *C* son dosis que están en estudio y, al comprobar que el resultado del Análisis de Varianza es significativo (se rechaza la H_0) se nos ocurre probar si el efecto de la dosis tradicional es distinto al efecto de las otras dosis experimentales, tomadas en su conjunto.
Esto es, la H_0 que queremos probar es que la media del grupo *A* es igual a la media de los grupos *B* y *C*.

$$H_0 : \mu_A = (\mu_B + \mu_C)/2$$

O bien: $H_0: \mu_1 - \frac{(\mu_2+\mu_3)}{2} = 0$

Si aplicamos la $CR_{Scheffé}$:

a) $n_i = 7 \quad I = 3 \quad \bar{Y}_1 = 11{,}7143 \quad \bar{Y}_2 = 15{,}4 \quad \bar{Y}_3 = 8{,}57$

b) El valor de la distribución F para un $\alpha = 0{,}05$ con 2 grados de libertad en el numerados y 18 grados de libertad en el denominador, es igual a 3,555.

c) $MC_{error} = 16{,}27$

d) Dado que sólo queremos realizar una comparación, sólo tenemos que establecer una combinación lineal para ver sus coeficientes (c_i). Dicha combinación será:

$$(2)\bar{Y}_1 + (-1)\bar{Y}_2 + (-1)\bar{Y}_3$$

Calculamos $CR_{Scheffé} = \sqrt{(2)(3{,}55)}\sqrt{(16{,}27)\left((2^2/7)+(1/7)+(1/7)\right)} = 9{,}95$

Esta es la diferencia mínima o rango crítico (***C****ritical* ***R****ange* de Scheffé), por encima de la cual (en valores absolutos) diremos que hay diferencias entre los grupos de medias que estemos comparando. Por lo tanto, realizamos la comparación:

$$\left|(2)(11{,}7143)+(-1)(15{,}43)+(-1)(8{,}57)\right| = 0{,}5714$$

e) Si comparamos 0,5714 < 9,95 concluimos que no podemos rechazar la hipótesis nula en este contraste *a posteriori* y, por lo tanto interpretamos, con un nivel del confianza del 95%, que el método tradicional no difiere de los otros dos tomados conjuntamente.

5.7. SUPUESTOS DEL ANÁLISIS DE VARIANZA

Tras ver en qué consiste el Análisis de Varianza, vamos a analizar ahora las condiciones que deben cumplir los datos para que pueda aplicarse el ANOVA. Estas son:

a) Independencia. Es decir, que las distintas muestras o grupos a comparar hayan sido obtenidas aleatoriamente. Esto implica que tanto las muestras como las observaciones deben ser independientes.
b) Normalidad de las distribuciones. Las muestras o grupos que comparamos deben proceder de poblaciones que se distribuyan normalmente en la variable estudiada.
c) Homogeneidad de las varianzas (homocedasticidad). Los grupos a comparar deben proceder de poblaciones que no difieran significativamente en sus varianzas en la variable estudiada.

Dado el alcance del temario no vamos a detenernos en el desarrollo de las distintas técnicas estadísticas que su suelen utilizar para probar que se cumplen estos supuestos. Para probar la independencia de las observaciones puede utilizarse, por ejemplo, el test de rachas; para probar la normalidad las pruebas más utilizadas son la Chi-cuadrado de Pearson, la prueba de Kolmogorov-Smirnov o la prueba de Lilliefors; y, para probar la homocedasticidad, se suele acudir a las pruebas de Cochran, de Bartlett o el test de Levene.

Aún insistiendo en la importancia de probar que se cumplen los supuestos, antes de proceder a realizar un Análisis de Varianza, resulta imposible (por la extensión del curso) desarrollarlos aquí. Los paquetes estadísticos para el cálculo del Análisis de Varianza comienzan probando los supuestos.

5.9.- RESUMEN

En este capítulo hemos iniciado el estudio del análisis de los diseños de más de dos muestras, comenzando por el caso más sencillo: el de un factor (variable independiente) con más de dos niveles o grupos, en el que cada nivel está integrado por un conjunto distinto de sujetos, asignados de manera aleatoria. Esta técnica de análisis se conoce como Análisis de Varianza (ANOVA)

- La técnica de análisis difiere de la seguida en el diseño de dos grupos, en el cual se contrasta la hipótesis nula de diferencias de medias, mediante la comparación directa de éstas. En el ANOVA, aunque la hipótesis estadística de partida es semejante "las medias de los grupos son iguales", el enfoque del análisis es muy diferente pues, como hemos visto, se hace a través del estudio de dos variabilidades como estimadores de la variabilidad total del sistema (de ahí su nombre).

- Una vía de estimación es a través de la variabilidad que se da en las puntuaciones individuales de cada grupo y otra es la variabilidad que se da entre los promedios generales de cada grupo. Si estas dos estimaciones son similares se deduce que no hay un efecto diferencial entre los grupos, pero si no lo son, la conclusión es que, además de la variabilidad intrínseca de un experimento, fruto del error asistemático, hay otra fuente de variabilidad fruto de las propias diferencias, estás sistemáticas, de los tratamientos entre sí.

- El estadístico F es el que, a la postre, determina si la comparación de estas variabilidades (denominadas Medias Cuadráticas) es o no estadísticamente significativa. Esta prueba es la que se conoce como prueba general o prueba *ómnibus*, y sirve para el contraste de la hipótesis nula general de igualdad de medias de las poblaciones objeto de estudio.

- Realizada la prueba general, si resulta significativa, se daría un paso más en la dirección de establecer entre qué grupos o combinaciones de grupos (dependiendo del tipo de contraste) se producen estas diferencias. Para estos contrastes hay una serie de estadísticos (Scheffé, entre otros) que informan de la significación estadística de las diferencias.

- También hemos visto el modelo teórico en que se basa el ANOVA, y los supuestos subyacentes: independencia; normalidad y homocedasticidad.

5.10.- EJERCICIOS DE AUTOEVALUACIÓN

5.10.1.- Enunciados

Se diseñó un experimento[5] para evaluar el efecto de diferentes tipos de entrenamiento para que los niños adquieran la habilidad del concepto "triángulo equilátero". En total participaron en la experiencia 40 niños de tres años de una escuela infantil, que fueron asignados aleatoriamente a cada uno de los siguientes cuatro grupos:

- Nivel a_1. Condición visual. Se mostraron 30 figuras de madera de forma consecutiva y la instrucción era mirarlos pero no tocarlos.
- Nivel a_2. Condición visual y motora. Los niños observan los bloques y además pueden jugar con ellos. Se les pide que realicen ejercicios táctiles específicos tales como desplazar el dedo índice por el perímetro de las piezas.
- Nivel a_3. Condición visual y verbal. Los niños miran los bloques y se les hace notar las diferencias en color, forma, tamaño y grosor.
- Nivel a_4. Los niños en este nivel no realizan actividad alguna.

Todas las actividades se desarrollaron de manera individual. El día después del entrenamiento, a cada niño se les enseñaba durante cinco segundos una figura y luego se le pedía que identificara esa pieza entre un conjunto de 7 piezas. La tarea se repitió seis veces usando diferentes figuras de referencia. Como variable dependiente se tomó el número de piezas correctamente identificadas. Los resultados se muestran en la siguiente tabla

a_1	a_2	a_3	a_4
0	2	2	1
1	3	4	0
3	4	5	2
1	2	3	1
1	1	2	1
2	1	1	2
2	2	3	1
1	2	3	0
1	3	2	1
2	4	4	3

[5] Los datos y el diseño del experimento están adaptados del libro ya referido de Kirk, R.E. *Experimental Designs. Procedure for de Behavioral Sciences.* Pág 206, que a su vez toma de una investigación de Nelson, G.K. (1976) *Concomitant effects of visual, motor, and verbal experiences in young children's concept development. Journal of Educational Psychology*, *68*, 466-473.

Preguntas

1. ¿Cuál es el resultado de las Media Cuadrática Intra-grupos ($MC_{S/A}$)?
    a. 1,0361
    b. 4,1536
    c. 37,300
2. ¿Cuál es el resultad de la Suma de Cuadrado Inter-grupos (SC_A)?
    a. 56,975
    b. 19,675
    c. 37,300
3. ¿Cuál es el resultado del estadístico de la prueba ómnibus?
    a. 6,330
    b. 2,866
    c. 1,314
4. Si queremos contrastar la hipótesis $\left(H_0: \mu_3 - \frac{\mu_1+\mu_2+\mu_4}{3} = 0\right)$, ¿cuál sería el valor de la diferencia crítica ($CR_{Scheffé}$) según la Prueba de Scheffé?
    a. 3,265
    b. 1,372
    c. 0,956

5.10.2.- Soluciones

Pregunta 1. a
Pregunta 2. b
Pregunta 3. a
Desarrollo del ANOVA

	a_1	a_2	a_3	a_4	Total
	0	2	2	1	
	1	3	4	0	
	3	4	5	2	
	1	2	3	1	
	1	1	2	1	
	2	1	1	2	
	2	2	3	1	
	1	2	3	0	
	1	3	2	1	
	2	4	4	3	
Suma	14	24	29	12	**79**
Media	1,4	2,4	2,9	1,2	**1,975**
Observ.	10	10	10	10	**40**

Tabla del ANOVA con las razones básicas ya calculadas (el lector debe desarrollar estos cálculos siguiendo el esquema ya presentado en el capítulo)

FV	Razones	SC	GL	MC	F	Prob.
A	[A] = 175,7	[A] - [T] = 19,675	3	6,558	6,330	0,001
S/A	[Y] = 213	[Y] - [A] = 37,3	36	1,036		
Total	[T]= 156,025	[Y] - [T] = 56,975	39			

Pregunta 4. a
Prueba de Scheffé
Coeficientes para la prueba

	a_1	a_2	a_3	a_4	
c	-1	-1	3	-1	
c^2	1	1	9	1	$\sum c^2 = 12$

$$\boldsymbol{CR_S = \sqrt{(I-1)\quad (F_{\alpha;v_1;v_2})\left(MS_{S/A}\sum\frac{c_i^2}{n_i}\right)}}$$

$$(4-1)F_{0,05;3;36} = (4-1)\times 2.86 = 8,58$$

$$CR_S = \sqrt{(8,58)\left(1,036\times\frac{12}{10}\right)} = 3,265$$