

## Capítulo 5: Diferencias que presenta una variable numérica entre varios grupos

En este capítulo vamos a generalizar las ideas del anterior, donde comparábamos las diferencias observadas en una variable en dos grupos, al caso de más grupos. De hecho, si aplicamos algunas de las técnicas que aquí veremos a sólo dos grupos, darían los mismos resultados que técnicas tratadas con anterioridad.

Como ilustración del tipo de problemas que podemos tratar, está el de comparar los resultados de 4 tratamientos. Uno podría ser un placebo aplicado a un grupo de control, y los otros tres podrían corresponder a sendas dosis de un medicamento. La variable respuesta sería una variable numérica, aunque también podría extenderse a variables ordinales.

Al igual que en el capítulo anterior consideraremos dos tipos de técnicas:

- Paramétricas: Cuando suponemos que las distribución de los datos de cada muestra tienen cierta distribución particular. La técnica que mostraremos se denomina ANOVA de una vía.
- No paramétricas: No es necesario suponer nada sobre la distribución de los datos. Interpretaremos la prueba de Kruskal-Wallis, que puede extenderse a variables ordinales.

Sea cual sea el contraste que utilicemos, la hipótesis nula traducirá la idea de que en los diferentes grupos se obtienen resultados similares (“*no efecto*”), y la hipótesis alternativa lo negará. La significación del contraste nos dará una idea de si las diferencias observadas en los diferentes grupos son imputables al azar (significación grande) o hay una diferencia intrínseca entre algunos grupos (significación pequeña).

### 5.1 Anova de un factor o una vía

Para usar el modelo ANOVA<sup>5</sup> debe poder suponerse que son válidas una serie de condiciones:

- La variabilidad de todas las muestras debe ser similar. Ésta es la condición más importante.
- Las muestras deben tener una distribución aproximadamente normal. Cierta alejamiento de esta hipótesis no es muy problemático.
- Los tamaños de las muestras no deben ser muy dispares. Esta condición en realidad no es estrictamente necesaria, y además es controlable al realizar un experimento. Pero, si se sospecha una cierta heterogeneidad en la variabilidad en los diferentes grupos (como es razonable esperar en la práctica), el tener grupos desequilibrados en tamaño sólo puede empeorar las cosas. Esta condición simplifica el poder interpretar los resultados.

Si se dan las dos primeras condiciones, el modelo ANOVA traduce la hipótesis nula de que en todos los grupos se obtienen valores similares de las variables por una condición equivalente: Que las medias (como parámetros) en los diferentes grupos son iguales.

---

<sup>5</sup> ANOVA es un acrónimo inglés de ANalysis OF VAriance.

### 5.1.1 En qué se basa el contraste ANOVA

Si suponemos que la hipótesis nula de la igualdad de medias en los diferentes grupos es cierta, podríamos decir que todas las observaciones pueden considerarse que provienen de un único grupo cuya media y variabilidad es la misma que la de cualquiera de los grupos por separado. Por tanto observamos que hay diferentes maneras de estimar la variabilidad en la población. De las discrepancias entre diferentes estimaciones de la variabilidad surge toda una familia de técnicas conocidas como análisis de la varianza, de las cuales el ANOVA de una vía no es más que el representante más simple.

La cuestión es que si alguno de los grupos presenta unos valores que en media se alejan del resto, esto se apreciará en el contraste como una fuente extra de variabilidad no explicable por el azar. La significación del contraste se calcula evaluando si esta variabilidad extra es muy grande con respecto a una variabilidad que sería de esperar si la hipótesis nula fuese cierta. Es por ello, que al realizar un contraste ANOVA siempre veremos varias fuentes de variabilidad. Los detalles son engorrosos, y en esta explicación resumida de la técnica nos limitaremos a tomar la decisión de rechazar o aceptar la hipótesis nula en función de la significación del contraste (que es fácilmente identificable).

Suele llevar a confusión el nombre, pues hace pensar que estamos contrastando igualdad de varianzas, pero en realidad esto es algo que se supone. Lo que contrastamos es la igualdad de medias.

### 5.1.2 Cómo se interpreta ANOVA

Si al realizar la prueba ANOVA se obtiene una significación baja (por ejemplo menor que 0.05) rechazaremos la hipótesis de que en todos los grupos las medias son iguales. La siguiente cuestión que aparece de modo natural en esta situación es la de identificar en qué grupos se han producido las diferencias. Básicamente tenemos las siguientes aproximaciones para abordar la cuestión:

- Cuando no tenemos una idea previa de en qué grupos eran de esperar las mayores diferencias, utilizamos los **contrastes no planeados** o **contrastes *post-hoc***.
- Antes de recoger datos es posible que tengamos algunas sospechas de dónde se deberían producir las diferencias. A esto se le denomina **comparaciones planeadas**. En SPSS lo encontramos pulsando el botón “contrastes” en la ventana para realizar el contraste ANOVA.

### 5.1.3 Contrastes no planeados o post-hoc

Bajo ese nombre encontramos múltiples técnicas. Éstas se consideran bastante conservadoras, en el sentido de que intentan reducir la posibilidad de errores de tipo I, a costa de aumentar la posibilidad de errores de tipo II. Dicho de otro modo, es probable que en situaciones donde realmente haya diferencias entre grupos, las pruebas *post-hoc* no lo detecten. Tienen que ser las diferencias entre grupos realmente grandes para poder ser reconocidas por estas pruebas.

Como hemos mencionado, dentro de la categoría de contrastes no planeados, hay muchas técnicas disponibles. Debe ser por algo. A continuación presentamos unas indicaciones que nos pueden ayudar a decidir cuál se adecua mejor a un estudio si estamos usando SPSS. Sólo indicaremos las más populares.

Una primera división entre estas pruebas es:

- **Tests de rangos:** Son aquellas que buscan identificar grupos homogéneos (medias parecidas).
- **Comparaciones múltiples:** Buscan establecer diferencias entre grupos basándose en diferencias dos a dos.

Esta división no es estricta pues hay técnicas que se pueden incluir en esas dos categorías a la vez.

- *Grupos equilibrados y varianzas similares:* Cuando todos los grupos tienen el mismo número de individuos, y podemos asumir que están igualmente dispersos.
  - Diferencia Honestamente Significativa de Tukey (**HSD de Tukey**). Se puede considerar a la vez como una técnica de comparaciones múltiples y a la vez de rangos. Es un test que se suele utilizar cuando se quiere comparar cada grupo con todos los demás y el número de grupos es alto (6 o más). Es una prueba conservadora (mantiene bajo el error de tipo I, sacrificando la capacidad de detectar diferencias existentes).
  - **Test de Scheffé:** Hace todas las comparaciones posibles. Por ejemplo, el primer grupo con respecto a cada uno de los restantes, pero también el primero con respecto al grupo formado por la unión de dos de los restantes, etc. Si sólo se desea comparar grupos dos a dos es preferible la prueba HSD de Tukey, pero si las comparaciones que buscamos son más complejas, preferimos esta prueba.
- *Grupos desequilibrados:* Cuando tenemos un número diferente de individuos en cada grupo puede interesarnos elegir alguna de las siguientes pruebas:
  - LSD de Fisher (sólo si hay 3 grupos),
  - T3 de Dunnett,
  - C de Dunnett,
  - Scheffé,
  - Games-Howell.
- *Varianzas desiguales:* Cuando la prueba de igualdad de varianzas (por ejemplo Levene) nos hace sospechar que las varianzas no son similares en todos los grupos, podemos considerar alguna de estas pruebas:
  - T2 de Tamhane
  - T3 de Dunnett,
  - C de Dunnett,
  - Scheffé,
  - Games-Howell.

### 5.1.4 Comparaciones planeadas

Son las que deberíamos hacer cuando *honestamente* tenemos una sospecha sobre el posible resultado del análisis, es decir entre qué grupos esperamos encontrar diferencias, y estas sospechas han sido formuladas *con anterioridad* a la recogida de datos.

Esto ocurre por ejemplo cuando a un grupo de control se le aplica un placebo, al segundo cierta cantidad de un medicamento experimental, y al tercero una cantidad mayor, pero que no esperamos en principio que mejore en mucho los resultados del segundo grupo. Posiblemente sospechemos que el grupo control va a obtener resultados muy diferentes al segundo y al tercero. No pensamos encontrar diferencias entre el segundo y el tercero. Este es el tipo de cuestiones que podemos resolver con las comparaciones planeadas.

Estas pruebas se realizan de la siguiente manera usando SPSS:

- En la ventana donde se realiza el contraste ANOVA (“Analizar – Comparar medias – ANOVA de un factor...”) pulsamos en el botón etiquetado “contrastos”. Allí marcamos la casilla “polinómicos”.
- Tomamos nota de un grupo (o más si pensamos que pueden ofrecer resultados parecidos) para compararlos con otro grupo (o más) que pensamos que ofrecerán resultados diferentes.
- A cada uno de los primeros les asociamos unos coeficientes positivos que reflejen la importancia de ese grupo, y a los grupos con los que queremos contrastarlos les asociamos unos coeficientes negativos. Esto se hace de modo que al sumarlos todos los coeficientes obtengamos un total de cero.
- Pulsando en “siguiente” podemos ir añadiendo todas las hipótesis que queramos.

Por ejemplo si pensamos que el primer grupo es diferente del segundo y el tercero, podemos asociarle al primer grupo un coeficiente 1 y -0.5 a cada uno de los otros dos; Así indicamos que queremos contrastar el primer grupo frente al conjunto formado por los dos restantes.

### 5.1.5 ¿Qué hacer si no se verifican las premisas del modelo ANOVA?

Hemos mencionado que para que el contraste ANOVA sea válido es necesario que podamos suponer una serie de propiedades sobre la distribución de la variable en los diferentes grupos.

La más importante es la de igualdad de las varianzas. Para ello podemos prestar atención al resultado de la prueba de **Levene**. Dado lo sensible que es el contraste ANOVA a la falta de este requisito, se acostumbra a rechazar la igualdad de varianzas con un nivel de significación superior al habitual (el 0.15=15% por ejemplo). En caso de que no se pueda admitir la igualdad de varianzas podemos usar las pruebas de **Welch** o de **Brown-Forsythe** para decidir sobre la igualdad de medias. En SPSS estas últimas, junto a la prueba de Levene, las encontramos al pulsar el botón “opciones...” que aparece en la misma ventana en la que se hace la prueba ANOVA.

Otra condición que necesitamos para la validez de la prueba ANOVA es la de la normalidad en cada uno de los grupos, aunque indicábamos que esta condición no es tan crítica como la de la igualdad de varianzas. Si la variable que medimos es un parámetro

biológico, con frecuencia tiene sentido hacer un cambio de variable logarítmico, que habitualmente normaliza las observaciones e iguala varianzas.

**Ejemplo:** Se realizó un experimento para comparar tres métodos de aprendizaje de lectura. Se asignó aleatoriamente los estudiantes a cada uno de los tres métodos. Cada método fue probado con 22 estudiantes (experimento equilibrado). Se evaluó mediante diferentes pruebas la capacidad de comprensión de los estudiantes, antes y después de recibir la instrucción. Las puntuaciones antes de recibir la enseñanza eran de media muy similares y aproximadamente normales. La prueba ANOVA con dichas puntuaciones no mostró diferencia estadísticamente significativa ( $p=0.45$ ). La prueba de Levene para la igualdad de varianzas confirmó que se podía asumir la igualdad de varianzas para las pruebas de lectura “antes” ( $p=0.738$ ). Es decir, los estudiantes de los tres grupos tenían habilidades medias similares, como sería de esperar al haber hecho la distribución aleatoriamente.

#### Prueba de homogeneidad de varianzas

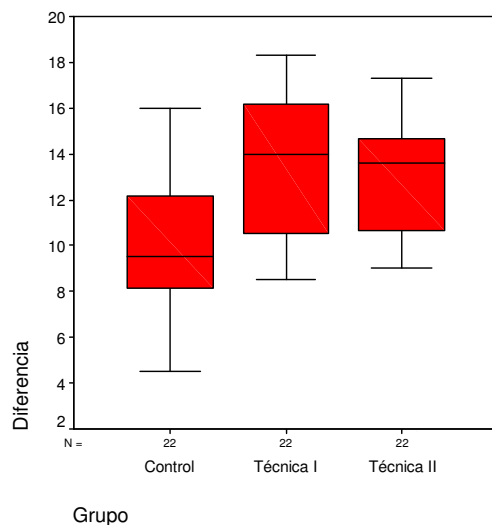
Antes

Estadístico de Levene	gl1	gl2	Sig.
,305	2	63	,738

#### ANOVA

Antes

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	7,826	2	3,913	,842	,436
Intra-grupos	292,739	63	4,647		
Total	300,564	65			



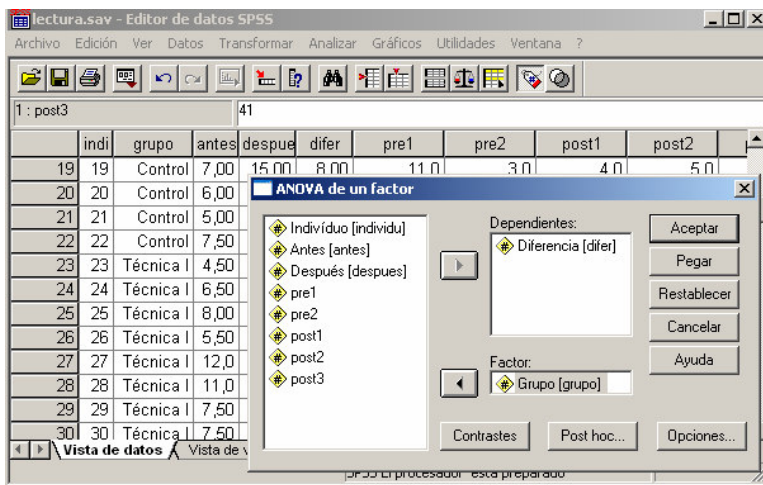
Queremos contrastar la hipótesis nula de que las tres técnicas ofrecen resultados medios similares, frente a la hipótesis alternativa de que esto no es así. Para ello podríamos considerar el resultado obtenido “después”. Pero esto estaría sin duda influido por la habilidad previa de cada estudiante. Para eliminar esta fuente de variabilidad podríamos considerar a cada estudiante como su propio control, y considerar una nueva variable que sería la diferencia entre “después” y “antes”. Ahora la hipótesis nula es que la variable “diferencia”, posee medias similares, frente a la hipótesis alternativa de que esto es

falso. Un análisis descriptivo exploratorio muestra que aparentemente el grupo control tiene resultados peores que los que siguieron la técnica I y II.

**Descriptivos**

Grupo		Estadístico	Error tít.	
Diferencia	Control	Media	9,8712	
		Mediana	9,5000	
		Varianza	7,156	
		Amplitud intercuartil	4,1667	
		Asimetría	,288	,491
		Curtosis	,121	,953
Técnica I	Técnica I	Media	13,5000	
		Mediana	14,0000	
		Varianza	9,381	
		Amplitud intercuartil	5,7500	
		Asimetría	-,161	,491
		Curtosis	-1,344	,953
Técnica II	Técnica II	Media	13,0909	
		Mediana	13,5833	
		Varianza	5,613	
		Amplitud intercuartil	4,1250	
		Asimetría	-,222	,491
		Curtosis	-,961	,953

Las diferencias tampoco se alejan demasiado de la normalidad y parecen similarmente dispersas. Parece que estamos entonces en condiciones de hacer una prueba ANOVA sobre la variable “diferencia” en los tres grupos. La prueba de Levene sobre igualdad de varianzas tiene una significación de  $p=0.251$ , lo que nos lo confirma. La prueba ANOVA resulta significativa siendo  $p$  un valor cercano a cero.



**Prueba de homogeneidad de varianzas**

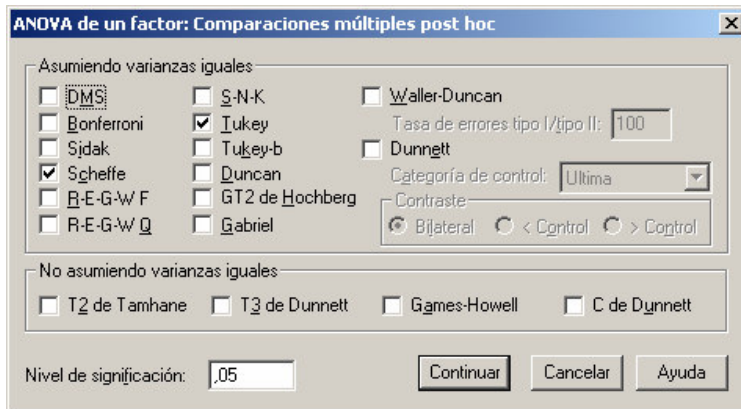
Diferencia			
Estadístico de Levene	gl1	gl2	Sig.
1,412	2	63	,251

**ANOVA**

Diferencia					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	173,814	2	86,907	11,771	,000
Intra-grupos	465,148	63	7,383		
Total	638,962	65			

Esto quiere decir que efectivamente aceptamos que hay diferencia estadísticamente significativa entre los resultados de las tres técnicas (las diferencias no han sido

causadas por el azar). Por tanto nos gustaría saber entre qué grupos se ha producido la diferencia. Una aproximación ingenua sería la de hacer directamente los contrastes no planeados, como si no tuviésemos ninguna hipótesis de qué grupo debe ser diferente al resto. Esto no es muy sensato, pues cuando hacemos un experimento no lo hacemos por que no tengamos nada mejor que hacer. Seguramente tendremos alguna sospecha. En cualquier caso, las pruebas de Scheffé y Tukey encuentran diferencias estadísticamente significativas para las combinaciones (grupo de control, técnica I), (grupo de control, técnica II), y no encuentra diferencias entre los grupos (técnica I, técnica II).



**Comparaciones múltiples**

Variable dependiente: Diferencia

	(I) Grupo	(J) Grupo	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
HSD de Tukey	Control	Técnica I	-3,62879*	,81927	,000	-5,5953	-1,6623
		Técnica II	-3,21970*	,81927	,001	-5,1862	-1,2532
	Técnica I	Control	3,62879*	,81927	,000	1,6623	5,5953
		Técnica II	,40909	,81927	,872	-1,5574	2,3756
	Técnica II	Control	3,21970*	,81927	,001	1,2532	5,1862
		Técnica I	-,40909	,81927	,872	-2,3756	1,5574
Scheffé	Control	Técnica I	-3,62879*	,81927	,000	-5,6828	-1,5748
		Técnica II	-3,21970*	,81927	,001	-5,2737	-1,1657
	Técnica I	Control	3,62879*	,81927	,000	1,5748	5,6828
		Técnica II	,40909	,81927	,883	-1,6449	2,4631
	Técnica II	Control	3,21970*	,81927	,001	1,1657	5,2737
		Técnica I	-,40909	,81927	,883	-2,4631	1,6449

\*. La diferencia entre las medias es significativa al nivel .05.

Observe que en todos los intervalos de confianza donde se ha obtenido diferencias estadísticamente significativas para parejas de grupos, se obtienen intervalos de confianza que no contienen el valor cero para la diferencia de medias. Esto es otra

manera de leer un contraste de hipótesis que permite obtener una interpretación clínica de los resultados de un experimento más interesante que la abstracta significación.

**Diferencia**

Grupo	N	Subconjunto para alfa = .05	
		1	2
HSD de Tukey <sup>a</sup>	Control	22	9,8712
	Técnica II	22	13,0909
	Técnica I	22	13,5000
	Sig.		1,000
Scheffé <sup>a</sup>	Control	22	9,8712
	Técnica II	22	13,0909
	Técnica I	22	13,5000
	Sig.		1,000

Se muestran las medias para los grupos en los subconjuntos homogéneos.

a. Usa el tamaño muestral de la media armónica = 22,000.

Junto a las pruebas anteriores, SPSS nos muestra una tabla formada por uniones de grupos que presentan valores de las medias parecidos.

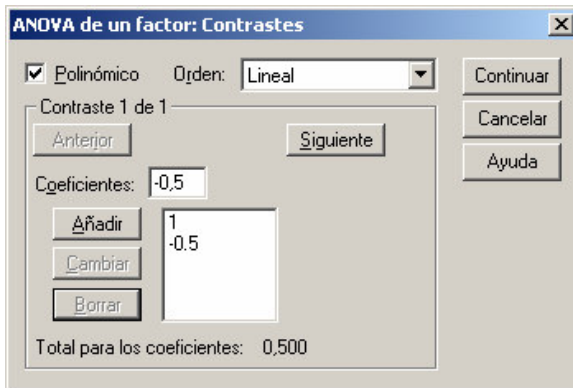
En este caso se aprecia que el grupo formado por los resultados de los individuos que han seguido las técnicas I y II formarían un grupo “homogéneo” (sus medias son muy similares, y un contraste de diferencia de medias resulta muy claramente muy *no significativo* ( $p=0,872$  para la prueba de Tukey y  $p=0,883$ ) para la prueba de Scheffe). Todo esto lo apreciamos en la columna etiquetada como ‘2’. La columna etiquetada como ‘1’, solo contiene un grupo, que por supuesto no presenta diferencias consigo mismo (faltaría más). SPSS lo muestra simplemente para que apreciemos la diferencia entre la media de ese grupo, con las cantidades que aparecen en la columna ‘2’.

**Coefficientes de los contrastes**

Contraste	Grupo		
	Control	Técnica I	Técnica II
1	1	-.5	-.5

Por último, recordemos que nuestra hipótesis de partida era bastante “fina”; Pensábamos (antes de recoger los datos), que el grupo control obtendría resultados diferentes a los que siguieron las técnicas I y II. Es legítimo entonces no hacer

comparaciones tan genéricas como las post-hoc, y hacer exactamente esa. Esto lo podemos realizar haciendo una comparación planeada. Asignamos coeficientes 1, -0.5 y -0.5 (juntos suman cero) respectivamente a los grupos Control, Técnica I y Técnica II, pulsando en el botón “contrastos” de la prueba ANOVA de un factor de SPSS. Los resultados resultan muy significativos para confirmar nuestra hipótesis (tanto suponiendo varianzas iguales como diferentes).



**Pruebas para los contrastes**

	Contraste	Valor del contraste	Error típico	t	gl	Sig. (bilateral)	
Diferencia	Asumiendo igualdad de varianzas	1	-3,4242	,70951	-4,826	63	,000
	No asumiendo	1	-3,4242	,70403	-4,864	42,556	,000

Es decir, concluimos que el grupo de control, presenta resultados diferentes al resto de los grupos (técnicas I y II).



## 5.2 Contraste no paramétrico de Kruskal-Wallis

Si tenemos dudas sobre la validez de las condiciones del ANOVA, o incluso si la variable respuesta es ordinal, la prueba no paramétrica de Kruskal-Wallis contrasta unas hipótesis análogas a la prueba ANOVA de un factor. No requiere condiciones de validez especiales.

El contraste de Kruskal-Wallis no contrasta que las medias sean iguales, sino simplemente si los valores obtenidos en los diferentes grupos son similares. Esto lo realiza de un modo sorprendentemente simple. Se ordenan todas las observaciones, de menor a mayor de todos los grupos. Si al ordenarse de esta manera se da la circunstancia de que muchas de las observaciones más pequeñas (o más grandes) pertenecen a un grupo sería una indicación de que los grupos no presentan valores similares. Es decir, si al ordenar las observaciones de menor a mayor, aparecen muy poco “mezcladas”, se rechaza la hipótesis nula.

**Ejemplo:** Continuamos el mismo ejemplo que usamos al explicar el contraste ANOVA. Como la prueba de Kruskal-Wallis no requiere ninguna condición de validez especial, podemos pasar a contrastar directamente si la variable diferencia presenta valores similares en los tres grupos. Esta prueba la encontramos en SPSS en el menú “Analizar - Pruebas no paramétricas - k muestras independientes”.



### Estadísticos de contraste<sup>a,b</sup>

	Diferencia
Chi-cuadrado	18,042
gl	2
Sig. asintót.	,000

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Grupo

La significación del contraste es prácticamente nula, es decir hay una gran evidencia estadística en contra de que los resultados obtenidos por las tres técnicas sean similares.

