Capítulo 11

Análisis de Varianza Simple (o con un factor), Factorial y Multivariable

Capítulo 11

Análisis de Varianza Simple (o con un factor), Factorial y Multivariable

1. Introducción

En capítulos precedentes hemos visto cómo estudiar la relación entre variables nominales, ordinales o de intervalo. Para las dos primeras partíamos del análisis de sus correspondientes tablas de contingencia; mientras que para las variables métricas evaluábamos su relación e intensidad a partir del modelo de regresión. Sin embargo, en investigación social en muchas ocasiones resulta imprescindible recurrir a la combinación de variables con niveles de medida distintos.

En estas circunstancias, las técnicas basadas en la **comparación de medias** son las más indicadas. Éstas, en líneas generales, calculan las medias de la variable dependiente para los grupos que forman las variables independientes: si se aprecian diferencias podremos concluir en que sí existe relación entre las variables relacionadas. Ahora bien, estas diferencias ¿son fruto del azar o son estadísticamente significativas? El **análisis de varianza** es uno de los contrastes que, como generalización de la prueba de la diferencia entre medias, nos va a permitir resolver la cuestión planteada. Esta técnica recibe este nombre porque la comparación entre las medias la efectúa en base al cálculo de la varianza entre tales medias.

En el análisis de varianza, a su vez, podemos diferenciar distintos tipos de análisis, a saber: (1) **análisis univariante de la varianza** y (2) **análisis multivariante de la varianza**. El primero de ellos se asocia con el análisis ANOVA (***AN****alysis* ***O****f* ***VA****riance*) y en él se valoran las diferencias entre grupos utilizando una única variable dependiente métrica. Por su parte, el análisis multivariante de la varianza se identifica con el análisis MANOVA (***M****ultivariate* ***AN****alysis* ***O****f* ***VA****riance)* y con él se analizan las diferencias entre grupos a partir de múltiples variables dependientes métricas. Los dos primeros apartados de este capítulo están dedicados a la exposición de las principales cuestiones que, desde la perspectiva de la estadística informática, resultan imprescindibles a la hora de ejecutar un análisis ANOVA posponiendo al tercer y último apartado del capítulo la exposición del análisis MANOVA.

El objetivo que persigue el análisis **ANOVA**, o **análisis simple de la varianza**, es el de determinar el efecto de una variable nominal u ordinal sobre una métrica. Un ejemplo puede ayudar a comprender el propósito del análisis que nos ocupa. Imaginemos que estamos interesados en saber el nivel de ingresos de la muestra entrevistada. Puesto que ésta es una pregunta que, o bien no se contesta, o bien se falsea (suscita muchas suspicacias), nosotros podemos pensar que este indicador puede ser inferido a partir del conocimiento del nivel de estudios de la población. Esto es, el nivel de ingresos (variable dependiente o explicada con un nivel de medición métrico) puede ser determinada a partir del nivel de estudios alcanzados por la población (variable independiente o explicativa con un nivel de medición ordinal). Cuando el objetivo es predecir una variable dependiente métrica a partir de una única variable independiente (o factor) estamos aplicando, específicamente, una **análisis de varianza univariado** (o simple) **con un factor**.

Ocurre, sin embargo, y siguiendo con el ejemplo expuesto, que según la profesión de la población ocupada entrevistada, su nivel de ingresos también variará. De ahí que podamos explicar la variable ingresos a partir de la consideración conjunta de, en este caso, dos variables (o factores) independientes cuyo nivel medición es nominal u ordinal. En el caso de trabajar con más de una variable independiente aplicamos, específicamente, un **análisis** (univariado o simple) **factorial de la varianza**.

Por último, podemos estar interesados en querer predecir a partir de la consideración del nivel de estudios y la profesión desempeñada en el momento de la encuesta no solo el nivel de ingresos sino también el número de hijos en la unidad familiar. En este último e hipotético caso, el análisis que específicamente aplicaríamos sería el **análisis multivariable de la varianza**.

Estos análisis: (1) el análisis de varianza con un factor; (2) el análisis factorial de varianza; y (3) el análisis multivariable de la varianza, se exponen a continuación. Los tres presentan, básicamente, el mismo esquema metodológico pero el hecho de que en el segundo consideremos más de un factor y en el tercero una segunda variable dependiente hace que, respectivamente, introduzcan ciertas particularidades que no podemos obviar.

A. ANÁLISIS DE VARIANZA CON UN FACTOR

El **análisis de varianza con un factor** analiza el comportamiento de una variable dependiente en las subpoblaciones o grupos establecidos por los valores de una única variable (o factor) independiente. Se aplica para contrastar la hipótesis nula de que las muestras proceden de subpoblaciones en las que la media de la variable dependiente es la misma; o lo que es lo mismo, no existen diferencias significativas entre las medias observadas. Las diferencias son debidas al azar y por tanto las distintas muestras proceden de una misma población. Matemáticamente lo anotaríamos como sigue:

$$H_o: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$$

Los supuestos estadísticos paramétricos que han de cumplir los datos para que se les pueda aplicar el análisis de varianza son tres: deben proceder de muestras aleatorias simples, debe existir normalidad en la distribución de los datos y las varianzas de las subpoblaciones deben ser iguales.

El análisis de varianza con un factor que ocupa la primera parte de este capítulo lo hemos estructurado en tres apartados, a saber:

1. La primera parte del análisis trata de establecer si la media de la variable dependiente difiere, o no, significativamente en función de la variable independiente o factor. El **estadístico F** nos permite contrastar la hipótesis nula de igualdad de medias en los grupos.

2. Ya hemos comentado que sólo tiene sentido aplicar el análisis de varianza cuando los datos cumplen los requisitos paramétricos arriba señalados. Por ello, es imprescindible contrastar, fundamentalmente, la homogeneidad o igualdad de las varianzas de la variable dependiente en los grupos (subpoblaciones) establecidos por los valores de la variable independiente. Para este caso se aplica la **prueba de Levene**. Si el resultado nos llevara a rechazar tal supuesto, deberemos someter a los datos a un **proceso de transformación**.

3. La última parte del análisis de varianza que exponemos tiene por finalidad determinar cuál o cuáles de los diferentes niveles del factor son los que difieren entre sí. Para ello cabe aplicar diferentes **pruebas a posteriori** (Post Hoc) o métodos para comparaciones múltiples.

2. Análisis de varianza con un factor: valoración del ajuste global

El análisis de varianza se basa en que la variabilidad total de la muestra puede descomponerse en la variabilidad debida a las diferencias entre grupos y la debida a la diferencia dentro de los grupos. A partir de este supuesto, el análisis de varianza proporciona, para contrastar la H_0 de igualdad de medias entre los grupos, el **estadístico F** (compara la variabilidad debida a las diferencias entre grupos con la debida a las diferencias dentro de los grupos). Cuanto mayor sea el valor de F y menor su significación, más probabilidad de que existan diferencias significativas entre los grupos. En consecuencia, si el *p-valor* asociado al estadístico es menor que el nivel de significación (normalmente, 0.05) rechazaremos la hipótesis nula de igualdad de medias. En nuestro ejemplo el *p-valor* es igual a 0.002. Éste se ha efectuado sobre la variable dependiente transformada (proceso que se describe en el punto siguiente). Previamente la aplicación de esta prueba sobre la variable dependiente original arrojó resultados negativos de ahí que tuviéramos que transformarla. En la sección de resultados sólo se muestra la aplicación de la prueba de levene para la variable una vez transformada (TD9).

3. Prueba de Levene

Para que el análisis de varianza tenga sentido deberemos, además de superar satisfactoriamente el resultado de la prueba de análisis de varianza propiamente, certificar que la varianza en cada una de las subpoblaciones o grupos es la misma; o lo que es lo mismo, que las varianzas no son heterogéneas (supuesto paramétrico).

La **prueba de Levene** contrasta la H_0 de homogeneidad de varianzas de la variable dependiente en los grupos o subpoblaciones de la variable independiente (ver el subcuadro **Estadísticos** del cuadro de diálogo **Explorar** en el menú de **Analizar**: **Estadísticos descriptivos** o/y el subcuadro **Opciones** en el menú de **Analizar: Comparación entre medias: ANOVA de un factor**). Si el *p-valor* asociado al estadístico de contraste es menor que el nivel de significación fijado (normalmente, 0.05) rechazaríamos la H_0 de igualdad de varianzas y, con ello, obviaríamos uno de los supuestos paramétricos en los que se basa este análisis. En consecuencia, carecería de sentido aplicar dicho análisis.

Si nos encontráramos ante esta situación, una posible solución es la de **transformar** los datos hasta conseguir la homogeneidad entre las varianzas (ver el cuadro de diálogo **Calcular variable** en el menú de **Transformar**). El SPSS contempla una serie de transformaciones que al aplicarlas nos permiten no sólo estabilizar las varianzas sino también conseguir la supuesta normalidad. Es importante insistir en que el análisis de varianza sólo podrá seguir con la variable dependiente transformada. En nuestro ejemplo, como ya hemos comentado, la variable D9 INCLINACIÓN POLÍTICA no superó, inicialmente, la prueba de homogeneidad de varianzas. Por ello ha sido necesario buscar una transformación con la que alcanzar el citado supuesto. En concreto, y a partir de los datos arrojados por el gráfico de dispersión por nivel y el indicador de potencia para la transformación (ambos figuran en la sección de resultados), la transformación con la que conseguimos igualar las varianzas ha sido el cuadrado; esto es, los valores de la variable TD9 serán igual al cuadrado de los valores originales de dicha variable.

4. Comparaciones múltiples o métodos Post Hoc

El análisis de varianza concluye, pues, rechazando o aceptando la hipótesis nula de igualdad de medias. De rechazarla, podremos afirmar que entre las medias existe una diferencia significativa pero, ¿esta diferencia afecta a todos los pares de medias?; o por el contrario, ¿sólo afecta a algunos de ellos?

La finalidad de las **pruebas a posteriori** o **post hoc** no es otra que la de comparar las medias para cada par de grupos para poder, así, identificar dónde se producen las diferencias significativas. Para ello contamos con una gran variedad de métodos de comparación múltiple. Se diferencian en el modo en el que ajustan el grado de significación obtenido. Todos los métodos comparan todos los grupos a la vez y los ordenan de forma ascendente.

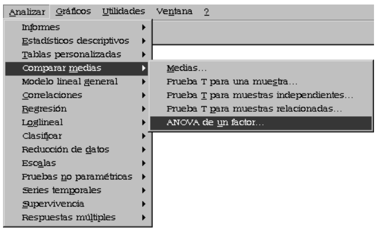


Figura 1

De todos, los métodos más utilizados son el **método de Scheffé** y el **método de Tukey.** El primero no exige que los tamaños muestrales sean iguales; el segundo sí. Sin embargo, el método de Tukey es mejor en la medida que detecta mejor las diferencias significativas. Por ello, y siempre y cuando los tamaños de las muestran no sean muy diferentes, podremos eliminar observaciones de los grupos de mayor tamaño hasta igualarlos y poder así aplicar el método de Tukey. Al aplicar el método se señala con un (*) los pares de medias en los grupos con diferencias significativas.

A continuación exponemos los distintos cuadros de diálogos y secuencias que debemos seguir para poder aplicar lo que de forma sintética hemos expuesto sobre el análisis de varianza con un factor. Nuestro interés se centra en saber si a partir de las preferencias del tipo de POLÍTICA a practicar respecto a la INMIGRACIÓN se puede inferir la INCLINACIÓN POLÍTICA de los encuestados.

5. Cuadro de Diálogo ANOVA de un factor

1er paso: El análisis de la varianza se solicita siguiendo la secuencia **Analizar: Comparar medias: ANOVA de un factor** (*figura 1*).

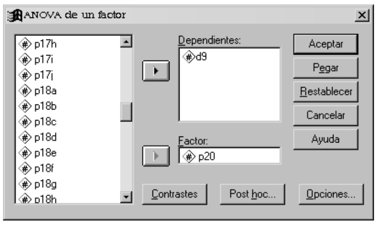


Figura 2

2º paso: Para poder empezar propiamente el análisis de varianza con un factor y siempre considerando nuestra hipótesis de trabajo, debemos especificar y seleccionar de la lista de variables que figuran en la matriz de datos una variable dependiente (cuantitativa o métrica) y una variable independiente (nominal u ordinal). Las variables seleccionadas, y a las que van a hacer referencia los resultados que presentamos, son: la P20, POLÍTICA INMIGRACIÓN (variable independiente y/o **Factor**); y la D9, INCLINACIÓN PÓLÍTICA (variable **Dependiente**) *(figura 2)*. Nótese que la variable dependiente seleccionada no es métrica. Esta selección responde a fines didácticos y expositivos y nos recuerdan que las técnicas a aplicar están supeditadas al tipo de datos disponibles y nunca los datos al servicio de aquellas.

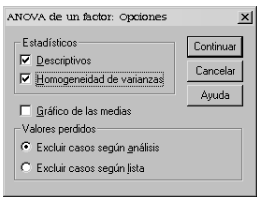


Figura 3

3er paso: Como ya hemos apuntado debemos comprobar, que las varianzas de los datos sobre los que estamos efectuando el análisis son iguales. Para ello en el botón de comando **Opciones** del cuadro de diálogo ANOVA de un factor podemos solicitar en

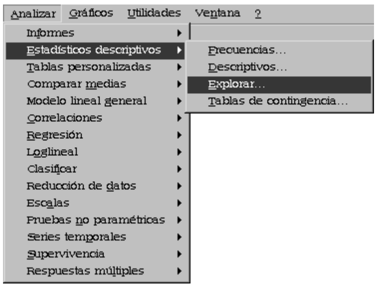


Figura 4

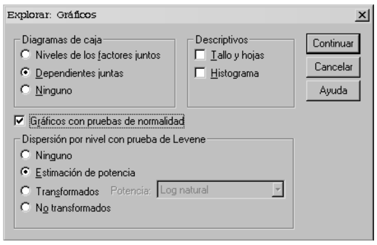


Figura 5

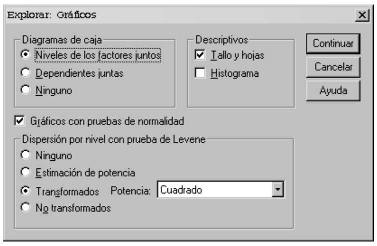


Figura 6

Figura 7

Estadísticos la tabla de los **Descriptivos** básicos y la prueba de Levene para contrastar la **Homogeneidad de varianzas** (*figura 3*).

En el ejemplo propuesto, el p-valor asociado al estadístico de contraste es menor a 0.05 por lo que debemos rechazar la hipótesis que contrastamos con la prueba de Levene, esto es, las varianzas no son iguales. Para igualarlas, y poder así aplicar el análisis de ANOVA con éxito, debemos transformar las variables.

4º paso: Para orientarnos sobre el tipo de transformación más adecuada podemos solicitar el gráfico de nivel y dispersión así como la estimación del poder de transformación para estabilizar la varianza de la variable dependiente. Esta información se obtiene siguiendo la siguiente secuencia: **Analizar: Estadísticos descriptivos: Explorar** (*figura 4*).

5.1. Cuadro de Diálogo Explorar

Al cliquear sobre el botón del comando **Gráficos** del cuadro de diálogo de Explorar podemos solicitar los **Gráficos con pruebas de normalidad** y la **Estimación de potencia** en la Dispersión por nivel con prueba de Levene (*figura 5*). La tabla donde aparece la prueba de homogeneidad de varianzas ofrece un *p-valor* asociado a la tabla inferior a 0.05. De la salida de este análisis cabe señalar que el *poder de transformación* estimado a partir de la recta de regresión ajustada a la nube de puntos es el dato que nos orienta sobre el tipo de transformación a ejecutar. Esta cifra habrá que redondearla al múltiple de un medio más próximo.

5º paso: Para transformar la variable dependiente habrá que seleccionar en Dispersión por nivel con prueba de Levene, en el mismo cuadro de diálogo, la opción **Transformados**. Al seleccionar esta opción se activa la ventana desplegable que recoge los distintos tipos de transformación. En nuestro caso la transfomación más indicada es al **Cuadrado** (*figura 6*).

6º paso: A continuación, volvemos al paso 4º pero ahora colocando la variable dependiente transformada (al cuadrado) (ver anexo II: transformaciones de las variable), y debemos solicitar la prueba de Levene pero ahora con los valores de la variable una vez transformada. La salida en este caso nos ofrece un *p-valor* asociado al estadístico de contraste ahora sí superior a 0.05 (*figura 7*).

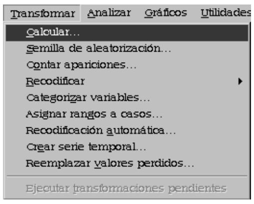


Figura 8

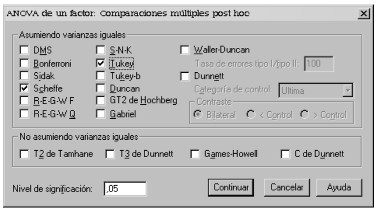


Figura 9

5.2. Cuadro de Diálogo Calcular

7° paso: Una vez superada esta prueba, el análisis seguirá tomado la variable transformada. Para ello, y puesto que ya sabemos qué transformación es la más indicada, podemos solicitarla a través del cuadro de diálogo **Calcular variable** del menú **Transformar** (*figura 8*) (ver Anexo II).

8° paso: Una vez que ya contamos con la variable transformada podemos finalizar el análisis de varianza. Por último, vamos a aplicar el método de comparación de medias más indicado al conjunto de datos con el que trabajamos con la finalidad de identificar diferencias significativas entre pares de medias y no en su conjunto. De todos los posibles métodos que figuran en este subcuadro aplicamos el de **Scheffé** y el de **Tukey**. Para solicitarlo aplicaremos de nuevo el análisis ANOVA con un factor pero, en esta ocasión, tomando como variable dependiente la variable dependiente transformada (TD9) y seleccionado dichos métodos en el subcuadro de diálogo **Comparaciones múltiples post hoc**, al cual se accede cliqueando el botón de comando **Post hoc** que aparece en la parte inferior del cuado principal de análisis de ANOVA de un factor (*figura 9*). El resultado de estos procedimientos determinará en una tabla qué medias de las diferentes categorías de la variable independiente son iguales o diferentes entre sí.

6. Resultados del análisis de varianza con un factor

Una vez finalizada la secuencia que acabamos de relacionar, obtenemos una relación de tablas que, en número, pueden verse ampliadas o reducidas, dependerá de si las varianzas de la variable dependiente son o no iguales. Con ellas vamos a determinar si las variables están relacionadas y si las podemos utilizar para explicar un determinado fenómeno.

- En primer lugar, se presenta la **prueba de Levene** para la variable transformada. Obviamente, y como paso previo, esta prueba fue aplicada en la variable original seleccionada no superándola (estos resultados han sido omitidos). Le sigue el **gráfico de dispersión por nivel** el cual nos orienta sobre el tipo de transformación a aplicar más indicada (en el que aparece el indicador de potencia para la transformación).

- A la prueba de levene y al gráfico de dispersión por nivel le sigue la tabla de **descriptivos básicos** asociados a las variables seleccionadas. En ella se establece, principalmente, el valor numérico de cada una de la medias. Es la primera aproximación, para poder determinar si las medias son diferentes o no.

A la tabla de descriptivos le sigue la tabla con los resultados del análisis **ANOVA** o de varianza con un factor. El estadístico F asociado (0.002) al estar por debajo del nivel de significación nos permite afirmar que existen diferencias entre las medias.

- La última tabla (en la que se aplican las **pruebas Post Hoc**) que nos recuerda que las diferencias significativas sólo se producen en algún par de grupos (aparecen señalados con un "*") y no en su conjunto. En concreto, las diferencias solo son significativas (y en los dos métodos) en el subgrupo delimitado por la combinación FAVORECE SU INTEGRACIÓN-NS/NC.

6.1. Pruebas de LEVENE. Variable dependiente transformada

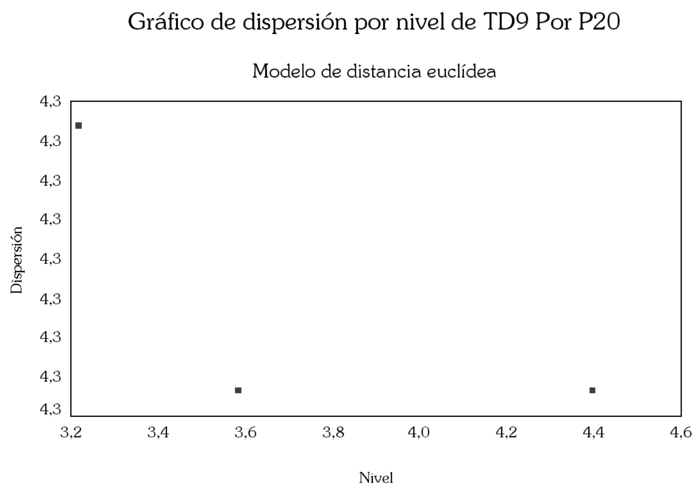
Pruebas de normalidad

	P20	Kolmogorov-Smirnov [a]		
		Estadístico	gl	Sig.
TD9	Favorecer su integración	,287	636	,000
	Favorecer su regreso	,305	458	,000
	ns/nc	,367	106	,000

a. *Corrección de la significación de Lilliefors*

Prueba de homogeneidad de la varianza

		Estadístico de Levene	gl1	gl2	Sig.
TD9	Basándose en la media	,884	2	1197	,413
	Basándose en la mediana.	,847	2	1197	,429
	Basándose en la mediana y con gl corregido	,847	2	851,252	,429
	Basándose en la media recortada	,728	2	1197	,483



* Gráfico de LN de dispersión por LN de nivel

Inclinación = -,048 Potencia para transformación = 1,048

6.2. ANÁLISIS DE VARIANZA UN FACTOR. Descriptivos básicos

Descriptivos

TD9

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95% Límite inferior	Límite superior	Mínimo	Máximo
Favorecer su integración	*636*	*39,7327*	*34,9810*	*1,3871*	*37,0089*	*42,4565*	*1,00*	*81,00*
Favorecer su regreso	*458*	*44,3275*	*34,0458*	*1,5909*	*41,2012*	*47,4538*	*1,00*	*81,00*
ns/nc	*106*	*51,8019*	*34,5004*	*3,3510*	*45,1575*	*58,4462*	*1,00*	*81,00*
Total	*1200*	*42,5525*	*34,7431*	*1,0029*	*40,5848*	*44,5202*	*1,00*	*81,00*

Prueba de homogeneidad de varianzas

TD9

Estadístico de Levene	gl1	gl2	Sig.
,884	*2*	*1197*	*,413*

6.3. Análisis de Varianza. Estadístico F

ANOVA

TD9

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	*15568,420*	*2*	*7784,210*	*6,508*	*,002*
Intra-grupos	*1431724,273*	*1197*	*1196,094*		
Total	*1447292,692*	*1199*			

6.4. Pruebas post hoc o comparaciones múltiples

Variable dependiente: TD9

	(I) P20	(J) P20	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
HSD de Tukey	**Favorecer su integración**	**Favorecer su regreso**	*-4,5948*	*2,1195*	*,077*	*-9,5622*	*,3726*
		ns/nc	*-12,0692**	*3,6283*	*,003*	*-20,5728*	*-3,5655*
	Favorecer su regreso	**Favorecer su integración**	*4,5948*	*2,1195*	*,077*	*-,3726*	*9,5622*
		ns/nc	*-7,4744*	*3,7277*	*,111*	*-16,2109*	*1,2621*
	ns/nc	**Favorecer su integración**	*12,0692**	*3,6283*	*,003*	*3,5655*	*20,5728*
		Favorecer su regreso	*7,4744*	*3,7277*	*,111*	*-1,2621*	*16,2109*
Scheffé	**Favorecer su integración**	**Favorecer su regreso**	*-4,5948*	*2,1195*	*,096*	*-9,7893*	*,5997*
		ns/nc	*-12,0692**	*3,6283*	*,004*	*-20,9615*	*-3,1769*
	Favorecer su regreso	**Favorecer su integración**	*4,5948*	*2,1195*	*,096*	*-,5997*	*9,7893*
		ns/nc	*-7,4744*	*3,7277*	*,134*	*-16,6102*	*1,6614*
	ns/nc	**Favorecer su integración**	*12,0692**	*3,6283*	*,004*	*3,1769*	*20,9615*
		Favorecer su regreso	*7,4744*	*3,7277*	*,134*	*-1,6614*	*16,6102*

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

B. ANÁLISIS FACTORIAL DE VARIANZA

El **análisis factorial de la varianza** analiza el comportamiento de la variable dependiente en las subpoblaciones o grupos establecidos por la combinación de los valores del conjunto de variables (o factores) independientes. Se aplica para contrastar la hipótesis nula de que las muestras proceden de subpoblaciones en las que la media de la variable dependiente es la misma; o lo que es lo mismo, no existen diferencias significativas entre las medias observadas. Las diferencias son debidas al azar y por tanto las distintas muestras proceden de una misma población. Matemáticamente lo anotaríamos como sigue:

$$H_o: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$$

Los supuestos estadísticos paramétricos que han de cumplir los datos para que se les pueda aplicar el análisis de varianza son tres: deben proceder de muestras aleatorias simples, debe existir normalidad en la distribución de los datos y las varianzas de las subpoblaciones deben ser iguales.

El análisis factorial de la varianza lo hemos estructurado en cuatro apartados, a saber:

1.- La primera parte del análisis trata de establecer si la media de la variable dependiente difiere, o no, significativamente en función de las variables independientes o factores. El **estadístico F**, nuevamente, nos permite contrastar la hipótesis nula de igualdad de medias en los grupos.

2. Ya hemos comentado que sólo tiene sentido aplicar el análisis de varianza cuando los datos cumplen los requisitos paramétricos arriba señalados. Por ello, es imprescindible contrastar, fundamentalmente, la homogeneidad o igualdad de las varianzas de la variable dependiente en los grupos (subpoblaciones) establecidos por los valores de la variable independiente. En estos casos se aplica la **prueba de Levene**. Si el resultado nos llevara a rechazar tal supuesto, deberemos someter a los datos a un **proceso de transformación**.

3. En tercer lugar, debemos identificar, tal y como ya hiciéramos en el análisis de varianza con un factor, en qué grupos se producen las diferencias significativas. Recordemos que estas diferencias son las que nos han conducido a rechazar la hipótesis nula. Para ello aplicaremos algunos de los **métodos Post Hoc**.

4. Por último, y si el análisis concluye en que rechazamos la hipótesis nula, nos restará por determinar si las diferencias que se aprecian entre las medias responde al efecto de las variables independientes consideradas de

forma individual (**efectos principales**) o/y responde al **efecto** provocado por la **interacción** de las variables independientes. Para resolver esta cuestión, además de los correspondientes contrastes de hipótesis, la representación **gráfica** de los resultados puede sernos de gran utilidad.

7. Análisis factorial de la varianza: valoración del ajuste global

El análisis de varianza se basa en que la variabilidad total de la muestra puede ser explicada por las diferencias entre grupos y por las diferencias que se producen en los grupos. A su vez, y esto es lo que le diferencia del análisis de varianza con un factor, la variabilidad explicada por el efecto de pertenecer a un grupo se descompone en: la variabilidad debida a los efectos de cada una de las variables independientes (efectos principales) y la debida al efecto de todas las posibles interacciones que se producen entre ellas.

Nuevamente a partir de este supuesto, el análisis de varianza proporciona, para contrastar la H_0 de igualdad de medias entre los grupos, el **estadístico F** (compara la variabilidad debida a las diferencias entre grupos con la debida a las diferencias dentro de los grupos). Cuanto mayor sea el valor de F y menor su significación, más probabilidad de que existan diferencias significativas entre los grupos. En consecuencia, si el *p-valor* asociado al estadístico es menor que el nivel de significación (normalmente, 0.05) rechazaremos la hipótesis nula de igualdad de medias.

8. Prueba de Levene

Ya expusimos en la sección anterior correspondiente que para que el análisis de varianza tenga sentido deberemos, además de superar satisfactoriamente el resultado de la prueba de análisis de varianza propiamente, certificar que la varianza en cada una de las subpoblaciones o grupos es la misma; es decir, que las varianzas no son heterogéneas (supuesto paramétrico).

La variable dependiente seleccionada es la misma que ya aplicamos en el análisis de varianza con un factor. No obstante, y teniendo en cuenta fines expositivos, en esta ocasión no la hemos transformada (D9): hemos supuesto que las varianzas son iguales. Ya sabemos que ante estas circunstancias no basta con "suponer" y que siempre deberemos aplicar la prueba de Levene para ratificar nuestra hipótesis. Dado que la variable original no superó la correspondiente prueba, nuestro análisis debería haber continuado con la transformación propuesta en el anterior ejemplo (TD9) cuyos valores, recordemos, eran iguales al cuadrado de los valores originales de dicha variable).

9. Comparaciones múltiples o métodos Post Hoc

Tal y como ya hiciéramos en el análisis de varianza de un factor y siempre y cuando la valoración del ajuste del modelo concluya en que rechazamos la hipótesis nula, nos queda determinar en qué grupos se produce una diferencia significativa. En este caso aplicaremos algunos de los **métodos de comparaciones múltiples** propuestos para este objetivo. De todos ellos, los que aplicaremos nuevamente serán los de **Scheffé** y **Tukey** (ambos han sido expuestos en el correspondiente apartado del primer punto de este capítulo). Las tablas correspondientes a estos análisis no han sido incluidas en la sección de resultados.

10. Análisis de la interacción entre los factores

En último lugar, y por lo que respecta al análisis factorial de la varianza, una vez rechaza la hipótesis nula de igualdad entre medias nos queda determinar qué efectos son los que influyen en el distinto comportamiento de la variable dependiente en los grupos establecidos. Para ello el análisis factorial de varianza determina, a partir de una serie de contrastes basados en el estadístico F, qué efectos (si los principales o/y el de las interacciones) son significativamente distintos de cero.

En al análisis factorial de la varianza cada uno de los efectos es contrastado con un estadístico F, en nuestro ejemplo al tener dos factores o variables independientes debemos contrastar: (1) la hipótesis nula de que el efecto de las interacciones de orden 2 es igual a 0: si el *p-valor* asociado al estadístico F es inferior a 0.05, esta interacción influye en que las medias sean diferentes; (2) la hipótesis nula de que el efecto del primer factor principal es igual a 0: si el *p-valor* asociado al estadístico F es inferior a 0.05, dicho factor influye en que las medias sean diferentes; y (3), se contrasta la hipótesis nula de que el efecto del segundo factor principal es nulo: si el *p-valor* asociado al estadístico F es inferior a 0.05, dicho factor influye en que las medias sean diferentes.

El término **interacción** alude al efecto conjunto de dos factores y es el efecto que debe ser examinado en primer lugar:

- Si el efecto interacción no fuera estadísticamente significativo, los efectos de los factores principales serán independientes. Por su parte, esta independencia nos advierte de que el efecto de un factor es el mismo para cada nivel del resto de factores y, en consecuencia, los efectos principales pueden interpretarse directamente.

- Si la interacción es significativa deberemos determinar el tipo de interacción. Las interacciones pueden ser consideradas como ordinal o disordinal (ver gráficos adjuntos): una **interacción ordinal** se produce cuando los efectos de un factor no son iguales para todos los niveles del resto de factores aunque la magnitud sea siempre de la misma dirección; mientras que en una **interacción disordinal** los efectos de un factor son positivos para algunos niveles y negativos para los otros niveles del resto de factores.

Las diferencias entre las interacciones son mejor descritas gráficamente. En concreto, el **gráfico de perfiles** (ver resultados) representa las medias de la variable dependiente en los grupos establecidos por la combinación de los valores de las variables independientes. No existe interacción cuando las líneas que representan las diferencias de un factor respecto al otro son paralelas (los efectos son constantes). En la interacción ordinal los efectos no son constantes (las líneas no son paralelas) apreciándose una variación en magnitud. Por último en las interacciones disordinales las líneas no solo no son paralelas sino que además entre los efectos se aprecia una variación tanto en magnitud como en dirección hasta el punto que las líneas terminan cruzándose.

En nuestro ejemplo puede apreciarse como el efecto interacción (línea RD8A*RD8 en la tabla prueba de efectos inter-sujetos) es estadísticamente significativo (0.002). Sin embargo, esta circunstancia no se reproduce en la significación de los dos efectos principales. En nuestro caso se aprecia como el efecto principal de la variable RD8 no es significativo (0.114), lo que se traduce en que las diferencias de medias no es explicada por este factor. Por su parte, el Gráfico de perfiles avala lo dicho.

Una vez expuestos los principales elementos a considerar en el análisis factorial de varianza, en las líneas que siguen se relaciona la secuencia de pasos a seguir cuando es el paquete estadístico SPSS la herramienta utilizada para llevar a cabo dicho análisis.

11. Cuadro de Diálogo del Análisis Factorial de la Varianza

Los intereses de nuestra investigación se centran ahora en determinar si a partir del conocimiento del tipo de RELIGIÓN practicada (D8) y de la INTENSIDAD DE LA PRÁCTICA RELIGIOSA (D8A) es posible saber la INCLINACIÓN POLÍTICA de la población sobre la que se ha extraído la muestra. Del mismo modo, también nos interesa saber si a partir del conocimiento, de forma individual, del tipo de RELIGIÓN practicada (D8), por un lado; y de la INTENSIDAD DE LA PRÁCTICA RELIGIOSA (D8A) de otro lado, es posible saber la INCLINACIÓN POLÍTICA del encuestado.

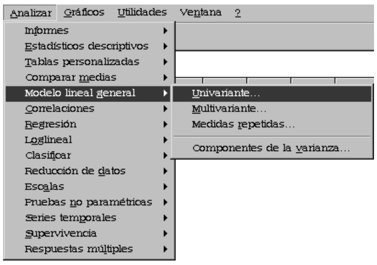


Figura 1

1er paso: Para acceder al cuadro de diálogo principal del análisis factorial de varianza seleccionamos de la barra del menú **Analizar: Modelo lineal general: Univariable** (*figura 1*).

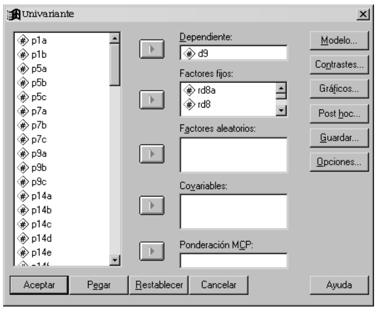


Figura 2

2º paso: Una vez en el cuadro de diálogo principal del análisis factorial de varianza Univariable, buscamos en la ventana de la izquierda las variables con las que vamos a trabajar. La variable D9, INCLINACIÓN POLÍTICA, la pasaremos al cuadro reservado para ubicar la variable **Dependiente**. Las variables RELIGIÓN e INTENSIDAD DE LA PRÁCTICA RELIGIOSA se colocarán en el cuadro destinado a las variables independientes o **Factores fijos** (*figura 2*). Nótese que ambas han sido previamente recodificadas por lo que las variables seleccionadas son las RD8 (1:católicos; 2: otra religión o ninguna; 9: NS/NC) y RD8A (1: no practican; 2: practican; 9: NS/NC).

Figura 3

3er paso: Antes de solicitar los estadísticos más indicados según nuestro intereses deberemos especificar el **Modelo** a aplicar. Para ello, deberemos cliquerar sobre el correspondiente botón de comando que aparece a la derecha del cuadro principal del cuadro de diálogo. Si nuestra elección ha sido **Modelo Factorial Completo** éste contemplará: el efecto de la intersección; todos los

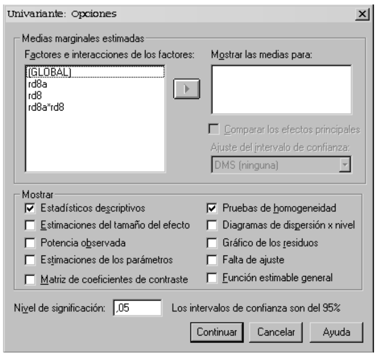


Figura 4

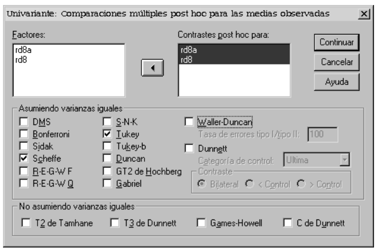


Figura 5

Figura 6

efectos principales de los factores y de las covariables; así como todas las intersecciones factor por factor. Una vez seleccionado el Modelo Factorial completo debemos seleccionar un método, de los cuatro posibles, para dividir la **Suma de los cuadros**. En nuestro caso concreto vamos a aplicar el **Tipo I** (*figura 3*).

4° paso: Con las variables seleccionadas debemos indicar aquellos parámetros que nos servirán para interpretar los resultados. Algunos de ellos se seleccionan en el subcuadro de diálogo de **Opciones**, al cual se accede cliqueando el botón de comando que aparece a la derecha del cuadro principal. Una vez en él pedimos que nos Muestre los **Estadísticos descriptivos** y las **Estimaciones del tamaño del efecto** (*figura 4*). En el supuesto de que no supiéramos si las varianzas de la variable dependiente son homogéneas pediríamos en este subcuadro la **Prueba de homogeneidad**. A partir de sus resultados arrojados por esta prueba repetiríamos los pasos ya descritos en el análisis de varianza de un factor.

5° paso: Con el objetivo de identificar en qué grupos se producen las diferencias significativas (en virtud de las cuales hemos podido rechazar la hipótesis nula) solicitaremos, como ya hiciéramos con el análisis de varianza de un factor, las pruebas de comparación múltiple o Post Hoc. Para ello cliqueamos sobre el botón de comando Post Hoc a través del cual accedemos al subcuadro **Comparaciones múltiples Post Hoc para las medias observadas**. De entre todos los métodos seleccionamos el de **Tukey** y el de **Scheffe** (*figura 5*). Los resultados han sido omitidos de la sección de resultados.

Por último, ya hemos comentado que con el fin de facilitar la interpretación respecto a los factores que influyen en el distinto comportamiento de la variable dependiente podemos solicitar el gráfico de las medias en los distintos grupos.

6° paso: En el subcuadro de **Gráficos de perfil**, al que se accede cliqueando en el botón que le da nombre y que se encuentra en la derecha del cuadro de diálogo principal del análisis factorial de la varianza, deberemos especificar sus parámetros. En el **Eje horizontal** siempre introduciremos la variable independiente con mayor número de categorías. El la casilla de **Líneas distintas** introduciremos la variable independiente con menor número de categorías. En nuestro caso y dado que ambas tienen el mismo número de categorías es indiferente (*figura 6*). Este gráfico especifica, siempre y cuando se produzca un cruce de líneas (hay interacción) si esa interacción es significativa y válida para explicar a la variable dependiente.

Finalizado el quinto paso, ya tenemos seleccionados todos los parámetros necesarios para realizar el análisis. De modo que después de cliquear **Continuar** (en el subcuadro de Gráfico en línea) y en el de **Aceptar** (en el cuadro de diálogo principal) el proceso se da por finalizado. Sólo resta analizar los resultados arrojados.

12. Resultados del Análisis Factorial de la Varianza

La salida e interpretación de los resultados sigue el mismo proceso que ya expusimos para el análisis de varianza con un factor con la salvedad de que hay que tener en cuenta que contamos con una variable independiente más y que, por ello, habrá que contemplar los efectos de las interacciones de las variables independientes sobre la dependiente (tabla con **pruebas de los efectos inter-grupos**). Para ello, recordemos, además de las correspondientes pruebas, el **gráfico de perfiles** puede sernos de gran utilidad. Este gráfico nos permite predecir los valores de la independiente a partir de los subgrupos delimitados por los factores. En el ejemplo que aportamos esta predicción no se puede concretar dado que la variable dependiente seleccionada, no es métrica, requisito imprescindible en el análisis de varianza.

En nuestro ejemplo y tras las pruebas practicadas parece confirmarse que la TENDENCIA POLÍTICA puede predecirse de forma individual exclusivamente a partir de la variable independiente INTENSIDAD DE LA PRÁCTICA RELIGIOSA así como a partir de la interacción conjunta de ambas variables independientes o factores.

12.1. Variables independientes o factores seleccionados

Factores inter-sujetos

		Etiqueta del valor	N
RD8A	**1**	*no practica*	*556*
	2	*practica*	*615*
RD8	**1**	*catolico*	*1107*
	2	*otra*	*64*

12.2. Estadísticos descriptivos

Variable dependiente: D9

RD8A	RD8	Media	Desv. típ.	N
no practica	**catolico**	*5,30*	*3,12*	*504*
	otra	*4,21*	*3,14*	*52*
	Total	*5,20*	*3,13*	*556*
practica	**catolico**	*6,26*	*2,81*	*603*
	otra	*7,58*	*2,75*	*12*
	Total	*6,29*	*2,81*	*615*
Total	**catolico**	*5,83*	*2,99*	*1107*
	otra	*4,84*	*3,32*	*64*
	Total	*5,77*	*3,02*	*1171*

12.3. Prueba de LEVENE

Contraste de Levene sobre la igualdad de las varianzas error[a]

Variable dependiente: D9

F	gl1	gl2	Sig.
10,862	*3*	*1167*	*,000*

Contrasta la hipótesis nula de que la varianza error de la variable dependiente es igual a lo largo de todos los grupos.

[a] *Diseño: Intercept+RD8A+RD8+RD8A * RD8*

12.4. Análisis de la varianza

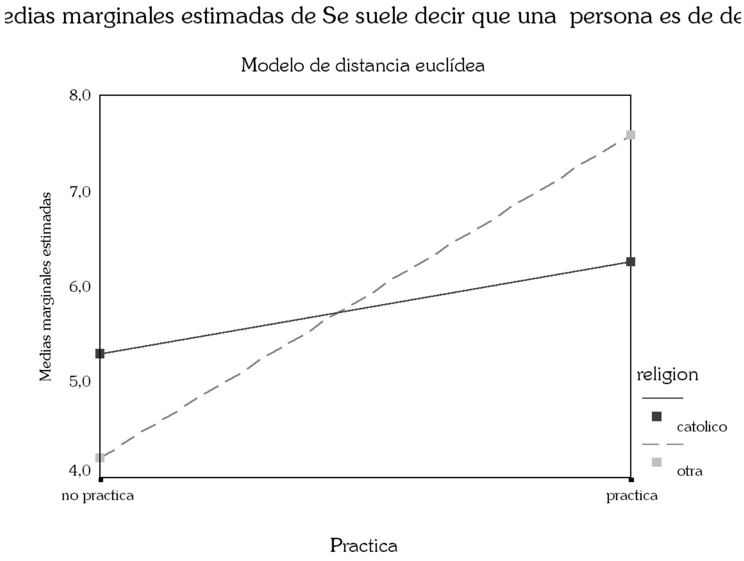
Pruebas de los efectos inter-sujetos

Variable dependiente: D9

Fuente	Suma de cuadrados tipo I	gl	Media cuadrática	F	Sig.
Modelo corregido	*421,378*[a]	*3*	*140,459*	*16,034*	*,000*
Intercept	*39012,879*	*1*	*39012,879*	*4453,602*	*,000*
RD8A	*344,624*	*1*	*344,624*	*39,341*	*,000*
RD8	*21,915*	*1*	*21,915*	*2,502*	*,114*
RD8A * RD8	*54,839*	*1*	*54,839*	*6,260*	*,012*
Error	*10222,744*	*1167*	*8,760*		
Total	*49657,000*	*1171*			
Total corregido	*10644,121*	*1170*			

[a] *R cuadrado = ,040 (R cuadrado corregido = ,037)*

12.5. Gráfico de Perfiles (Interacción).



C. ANÁLISIS MULTIVARIABLE DE LA VARIANZA

El **análisis multivariable de la varianza** es una extensión del análisis de la varianza (ANOVA) en donde se tiene en cuenta más de una variable de criterio o dependiente. Es, pues, una técnica de dependencia que analiza el comportamiento de un conjunto de variables métricas dependientes en las subpoblaciones o grupos establecidos por la combinación de los valores del conjunto de variables categóricas (o factores) independientes. El ANOVA y el MANOVA, como procedimientos de inferencia estadística, se aplican para contrastar la significación estadística de las diferencias entre grupos. Mientras que en ANOVA la hipótesis nula contrastada es la igualdad de las medias de la variable dependiente entre los grupos (se contrasta la igualdad entre los grupos de una única variable dependiente); en el MANOVA la hipótesis nula contrastada es la igualdad de un conjunto (**vector**) de medias de las variables dependientes entre los grupos delimitados por los valores de la/s variable/s (factor/es) independientes (se contrasta la igualdad de un valor teórico el cual combina óptimamente las medidas dependientes múltiples dentro de un valor único que maximiza las diferencias entre los grupos). El MANOVA contrasta la hipótesis nula de que los vectores de las medias de todos los grupos son iguales y/o provienen de la misma población. Matemáticamente lo anotaríamos como sigue:

$$H_o: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$$

Los supuestos estadísticos paramétricos que han de cumplir los datos para que se les pueda aplicar el análisis de varianza son tres: deben proceder de muestras aleatorias simples, debe existir normalidad en la distribución de los datos y las varianzas de las subpoblaciones deben ser iguales.

El análisis multivariante de la varianza, con el que finalizamos la exposición de este capítulo, lo hemos estructurado en torno a cinco apartados, a saber:

1. La primera parte del análisis trata de valorar la significación estadística de las diferencias multivariantes entre los grupos. Para este fin nos encontramos con cuatro estadísticos a partir de los cuales contrastar la hipótesis nula de igualdad de vectores de medias: **mrc de Roy**, **lambda de Wilks**, **criterio de Pillai** y la **traza de Hotelling**.

2. Ya hemos comentado que sólo tiene sentido aplicar el análisis de varianza cuando los datos cumplen los requisitos paramétricos arriba señalados. Por ello, es imprescindible contrastar, fundamentalmente, la homogeneidad o igualdad de las varianzas de las variables dependientes en los grupos (subpoblaciones) establecidos por los valores de la variable independiente. Para este caso se aplica la **prueba de Levene** en las dos variables dependientes seleccionadas. Si el resultado nos llevara a rechazar tal supuesto, deberemos someter a los datos a un **proceso de transformación**.

3. Si el análisis concluye en que rechazamos la hipótesis nula (todos los criterios se sitúan por debajo de 0.05), esto es, que existen diferencias entre entre los grupos, deberemos analizar si las diferencias que se producen entre vectores son explicadas por las diferencias que se producen entre las medias de alguna de las variables dependientes. Para contrastar si las diferencias entre los vectores de medias no son debidas a las diferencias entre

las medias de una de las variables dependientes en particular deberemos realizar un **análisis univariante sobre cada una de las variables dependientes**. Éste es el punto de la investigación que lo diferencia del análisis factorial de la varianza.

4. En cuarto lugar, debemos identificar, tal y como ya hiciéramos en el análisis de varianza con un factor y en el análisis factorial de la varianza, en qué grupos se producen las diferencias significativas. Recordemos que estas diferencias son las que nos han conducido a rechazar la hipótesis nula. Para ello **aplicaremos algunos de los métodos Post Hot** (de Sheffe y de Tuhey). Cada uno de estos métodos indican qué comparaciones entre los grupos presentan diferencias significativas. La aplicación de estos métodos proporciona los contrastes para cada combinación de grupos.

5. Por último, y si el análisis concluye en que rechazamos la hipótesis nula, nos quedará por **determinar** si las diferencias que se aprecian entre las medias responde al efecto de las variables independientes consideradas de forma individual (**efectos principales)** o/y responde al efecto provocado por la **interacción de las variables independientes**. Para resolver esta cuestión, además de los correspondientes contrastes de hipótesis, la representación gráfica de los resultados puede sernos de gran utilidad.

13. Análisis multivariante de la varianza: valoración del ajuste global

El análisis multivariante de la varianza con n factores se basa en que la variabilidad total de la muestra puede descomponerse en la variabilidad debida a las diferencias entre grupos y la debida a las diferencias dentro de los grupos. A su vez, y tal y como ya sucediera con el análisis factorial de varianza, la variabilidad explicada por el efecto de pertenecer a un grupo se descompone en: la variabilidad debida a los efectos de cada una de las variables independientes (efectos principales) y la debida al efecto de todas las posibles interacciones que se producen entre ellas.

A partir de este supuesto, el análisis multivariante de varianza proporciona, varios criterios con los que valorar las diferencias multivariantes entre los grupos (valorar las diferencias entre dimensiones de las variables dependientes o contrastar la hipótesis nula de igualdad de vectores de medias):

- **La raíz máxima (mrc) de Roy**. Mide las diferencias solamente sobre la primera raíz canónica (o función discriminante). El contraste mrc de Roy es el más indicado cuando las variables dependientes están fuertemente interrelacionadas en una sola dimensión aunque es la medida que se ve más afectada por el incumplimiento de los supuestos paramétricos.

- **Lambda de Wilks**. A diferencia del estadístico mrc que está basado en la primera (mayor) raíz característica, el lambda de Wilks considera todas las raíces características pues compara si los grupos son de algún modo diferentes sin estar afectados por el hecho de que los grupos difieran en al menos una combinación lineal de las variables dependientes. La significación del lambda de Wilks se realiza, como ya hemos expuesto, transformándolo en un **estadístico F** (compara la variabilidad debida a las diferencias entre grupos con la debida a las diferencias dentro de los grupos). Cuanto mayor sea el valor de F y menor su significación, más probabilidad de que existan diferencias significativas entre los grupos.

- Por último, la **traza de Pillai** y la **traza de Hotelling** son otras medias muy utilizadas en el análisis MANOVA. Ambas son muy similares al lambda de Wilks pues utilizan todas las raíces características y se aproxima a ellas a partir de un estadístico F.

El SPSS incluye estos cuatro criterios. Si los *p-valores* asociados a cada uno de los estadísticos son menores que el nivel de significación (normalmente, 0.05) se rechazará la hipótesis nula de igualdad de vectores de medias. En nuestro ejemplo no supera el ajuste la variable RD8 (ver tabla contrastes univariados).

14. Prueba de Levene

Ya hemos visto en los dos apartados anteriores y en sus correspondientes secciones, que para que el análisis de varianza tenga sentido deberemos, además de superar satisfactoriamente el resultado de la prueba de análisis de varianza propiamente, certificar que la varianza en cada una de las subpoblaciones o grupos es la misma; esto es, que las varianzas no son heterogéneas (supuesto paramétrico).

En el ejemplo que recogemos se puede apreciar como las variables dependientes seleccionadas no superan el correspondiente contraste de Levene. Pese a ello, y puesto que nuestro objetivo prioritario es el expositivo, el análisis continúa con las mismas. No obstante, debemos recordar que el análisis, en contextos de investigación real, sólo podría desarrollarse en él los términos que siguen si los datos de las variables dependientes no cumplieran el requisito paramétrico de igualdad de varianzas.

15. Análisis univariante de las variables dependientes

Los estadísticos Traza de Pillai, Traza de Hotelling, Lambda de Wilks y Raíz máxima de Roy nos permiten, siempre y cuando se sitúen por debajo del nivel de significación (normalmente 0.05), afirmar que existe un comportamiento diferente en los distintos grupos formados a partir de los valores de la/a variable/s independiente/s o factor/es. Sin embargo, cabe la posibilidad de que las diferencias que se aprecian entre los vectores se produjeran porque existen diferencias entre las medias de una de las dos variables. Para descartar tal posibilidad es necesario que, junto con el análisis multivariable propiamente, se ejecute un análisis univariante de cada una de las variables dependientes. Los resultados de este análisis no han sido incluidos en la sección correspondiente.

Si el *p-valor* asociado al estadístico F es menor que el nivel de significación (normalmente 0.05) para cada una de las variables dependientes sometidas a análisis, podremos concluir en que las diferencias entre los vectores de medias no son debidas a las diferencias entre medias de alguna de las variables dependientes. Rechaza, pues, la hipótesis nula, el proceso analítico concluye con dos secuencias ya contempladas en el análisis factorial de la varianza.

16. Comparaciones múltiples o métodos Post Hoc

Tal y como ya hiciéramos en el análisis de varianza de un factor y análisis factorial de la varianza y, siempre y cuando la valoración del ajuste del modelo global concluya en que rechazamos la hipótesis nula, nos queda determinar en qué grupos se produce una diferencia significativa. En este caso aplicaremos algunos de los **métodos de comparaciones múltiples** propuestos para este objetivo. De todos ellos, los que aplicaremos nuevamente serán los de **Scheffé** y **Tukey** (ambos han sido expuestos en el correspondiente apartado del primer punto de este capítulo). Los resultados de este análisis no han sido incluidos en la sección correspondiente.

17. Análisis de la interacción entre los factores

Si se rechaza la hipótesis nula de igualdad de vectores de medias nos queda determinar qué efectos son los que influyen en el distinto comportamiento de las variables dependientes en los grupos establecidos por la combinación de los valores del conjunto de variables independientes. Para ello, el análisis multivariante de varianza determina, a partir de una serie de contrastes basados en el estadístico F, qué efectos (si los principales o/y el de las interacciones) son significativamente distintos de cero.

En al análisis multivariante de la varianza cada uno de los efectos es contrastado con un estadístico F (ver tabla Pruebas de los efectos intra-grupos). En nuestro ejemplo al tener dos variables dependientes y dos variables o factores independientes debemos contrastar: en primer lugar, contrastaremos la hipótesis nula de que los vectores de medias son iguales en los grupos establecidos por la combinación de los valores de las q-1 primeras variables independientes: si el *p-valor* asociado a los estadísticos Traza de Pillai, Traza de Hotelling, Lambda de Wilks y Raíz máxima de Roy son inferiores a 0.05, podremos rechazar la hipótesis nula. Este contraste se repite para cada uno de los posibles subconjuntos formados por q-1 variables independientes. En el supuesto caso de que para un conjunto de variables independientes se rechaza la hipótesis nula, deberemos realizar un contraste univariante sobre cada una de las variables dependientes de forma individual.

Por último se plantea la hipótesis nula de que los vectores de medias son iguales en cada uno de los grupos establecidos por por los valores de la primera variable independiente: si el *p-valor* asociado a los estadísticos Traza de Pillai, Traza de Hotelling, Lambda de Wilks y Raíz máxima de Roy son inferiores a 0.05, podremos rechazar la hipótesis nula; esto es, el factor analizado influye en que los vectores de medias sean diferentes. Este contraste se repite para cada una de las variables independientes que hayan participado en la definición del modelo. En el supuesto caso de que para alguna de las variables independientes se rechaza la hipótesis nula, deberemos realizar un contraste univariante sobre cada una de las variables dependientes de forma individual.

En nuestro ejemplo puede apreciarse que el efecto principal de la variable RD8A es decisivo para explicar las variables dependientes pero que la interacción de los factores principales solo es significativa para la variable dependiente INCLINACIÓN POLÍTICA. Los correspondientes gráficos de perfil que se incluyen en la sección de resultados puede ayudar a entender lo expuesto.

18. Cuadro de Diálogo del análisis multivariante factorial

Los intereses de nuestra investigación se centran ahora en determinar si a partir del conocimiento de la RELIGIÓN practicada (D8) así como de la INTENSIDAD DE LA PRÁCTICA RELIGIOSA (D8A), es posible saber junto a la TENDENCIA POLÍTICA el ESTADO CIVIL (D3) de la población sobre la que se ha extraído la muestra. También estamos interesados en saber si a partir del conocimiento, de forma asilada, de la RELIGIÓN practicada, por un lado, y de la INTENSIDAD DE LA PRÁCTICA RELIGIOSA, es posible saber la TENDENCIA POLÍTICA y el ESTADO CIVIL de la población sobre la cual se ha extraído la muestra. Nótese que las variables dependientes seleccionadas no son variables métricas, condición imprescindible para la aplicación de esta técnica. Su elección responde a fines expositivos.

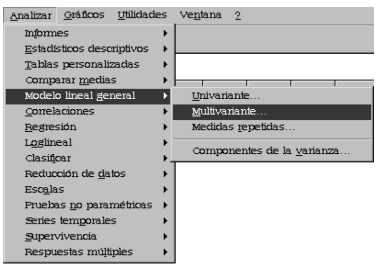


Figura 1

1er paso: Para acceder al cuadro de diálogo principal del análisis multivariable de varianza seleccionamos de la barra del menú **Analizar: Modelo lineal general: Multivariante** (*figura 1*).

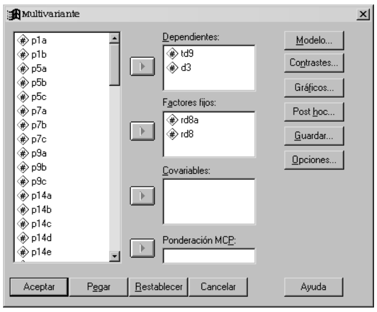


Figura 2

Figura 3

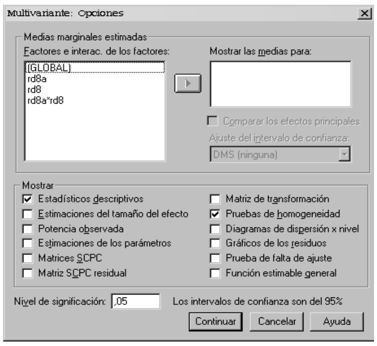


Figura 4

2º paso: Una vez en el cuadro de diálogo principal del análisis multivariable de la varianza, buscamos en el cuadro de la izquierda las variables con las que vamos a trabajar. Las variables TD9, TENDENCIA POLÍTICA y D3, ESTADO CIVIL, las seleccionamos y pasamos a la ventana reservado para ubicar las variables **Dependientes**. Las variables D8, RELIGIÓN y D8A, INTENSIDAD DE LA PRÁCTICA RELIGIOSA se colocarán en el cuadro destinado a las variables independientes o **Factores fijos** (*figura 2*). Nótese que las dos variables independientes han sido previamente recodificadas por lo que las variables seleccionadas son las RD8 (1:católicos; 2: otra religión o ninguna; 9: NS/NC) y RD8A (1: no practican; 2: practican; 9: NS/NC).

3er paso: Antes de solicitar los estadísticos más indicados según nuestro intereses deberemos especificar el **Modelo** a aplicar. Para ello, deberemos cliquerar sobre el correspondiente botón de comando que aparece a la derecha del cuadro principal del cuadro de diálogo. Si nuestra elección ha sido **Modelo Factorial Completo** éste contemplará: el efecto de la intersección; todos los efectos principales de los factores y de las covariables; así como todas las intersecciones factor por factor. Una vez seleccionado el Modelo Factorial completo debemos seleccionar un método, de los cuatro posibles, para dividir la **Suma de los cuadros**. En nuestro caso concreto vamos a aplicar el **Tipo I** (*figura 3*).

4º paso: Con las variables seleccionadas debemos indicar aquellos parámetros que nos servirán para interpretar los resultados. Algunos de ellos se seleccionan en el subcuadro de diálogo de **Opciones**, al cual se accede cliqueando el botón de comando que aparece a la derecha del cuadro principal. Una vez en él pedimos que nos Muestre los **Estadísticos descriptivos** y las **Estimaciones del tamaño del efecto** (*figura 4*). En el supuesto de que no supiéramos si las varianzas de la variable dependiente son homogéneas pediríamos en este subcuadro la **Prueba de homogeneidad**. A partir de sus resultados repetiremos los pasos ya descritos en el análisis de varianza de un factor.

5º paso: Con el objetivo de identificar en qué grupos se producen las diferencias significativas (en virtud de las cuáles hemos podido rechazar la hipótesis nula) solicitaremos, como ya hiciéramos con el análisis de varianza de un factor, las pruebas de comparación múltiple o Post Hoc. Para ello cliqueamos sobre el botón de comando Post Hoc a través del cual accedemos al subcuadro **Comparaciones múltiples Post Hoc para las medias**

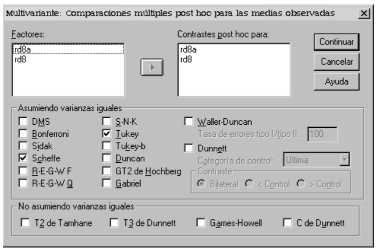


Figura 5

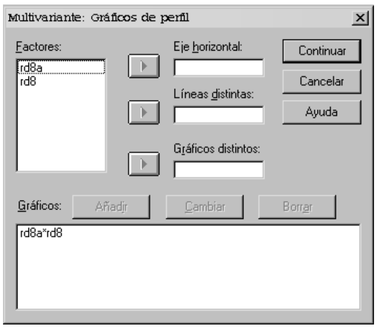


Figura 6

observadas. De entre todos los métodos seleccionamos el de **Tukey** y el de **Scheffe** (*figura 5*).

Por último, ya hemos comentado que con el fin de facilitar la interpretación respecto a los factores que influyen en el distinto comportamiento de la variable dependiente podemos solicitar el gráfico de las medias en los distintos grupos.

6° paso: En el subcuadro de **Gráficos de perfil**, al que se accede cliqueando en el botón que le da nombre y que se encuentra en la derecha del cuadro de diálogo principal del análisis factorial de la varianza, deberemos especificar sus parámetros. En el **Eje horizontal** siempre introduciremos la variable independiente con mayor número de categorías. El la casilla de **Líneas distintas** introduciremos la variable independiente con menor número de categorías. En nuestro caso y dado que ambas tienen el mismo número de categorías es indiferente (*figura 6*). Este gráfico especifica, siempre y cuando se produzca un cruce de líneas (hay interacción) si esa interacción es significativa y válida para explicar a la variable dependiente.

Finalizado el quinto paso, ya tenemos seleccionados todos los parámetros necesarios para realizar el análisis. De modo que después de cliquear **Continuar** (en el subcuadro de Gráfico en línea) y en el de **Aceptar** (en el cuadro de diálogo principal) el proceso se da por finalizado. Sólo resta analizar los resultados arrojados.

19. Resultados del análisis multivariable de la varianza

La salida e interpretación de los resultados sigue el mismo proceso que ya expusimos para el análisis de varianza con un factor y análisis factorial de la varianza. Es importante el análisis de los efectos (tabla con **pruebas de los efectos inter-grupos**). Para ello, recordemos, además de las correspondientes pruebas, el **gráfico de perfiles** puede sernos de gran utilidad. Este gráfico nos permite predecir los valores de la independiente a partir de los subgrupos delimitados por los factores. En el ejemplo que aportamos esta predicción no se puede concretar dado que la variable dependiente seleccionada no es métrica, requisito imprescindible en el análisis de varianza.

En nuestro ejemplo, y tras las pruebas practicadas, podemos apreciar que: (1) el efecto de la interacción entre las dos variables independientes sólo es significativo para la variable TENDENCIA POLÍTICA, pero apenas dicen nada sobre el ESTADO CIVIL de la población; y (2), sólo el efecto principal de la variable INTENSIDAD DE LA PRÁCTICA RELIGIOSA explicaría las variables dependientes.

19.1. Variables independientes o factores seleccionados

Factores inter-sujetos

		Etiqueta del valor	N
RD8A	**1**	*no practica*	*556*
	2	*practica*	*615*
RD8	**1**	*catolico*	*1107*
	2	*otra*	*64*

19.2. Estadísticos descriptivos

	RD8A	RD8	Media	Desv. típ.	N
TD9	**no practica**	**catolico**	*37,83*	*35,27*	*504*
		otra	*27,40*	*34,61*	*52*
		Total	*36,85*	*35,31*	*556*
	practica	**catolico**	*47,09*	*33,32*	*603*
		otra	*64,42*	*31,02*	*12*
		Total	*47,43*	*33,34*	*615*
	Total	**catolico**	*42,87*	*34,52*	*1107*
		otra	*34,34*	*36,74*	*64*
		Total	*42,41*	*34,68*	*1171*
D3	**no practica**	**catolico**	*1,93*	*1,16*	*504*
		otra	*1,65*	*,65*	*52*
		Total	*1,90*	*1,13*	*556*
	practica	**catolico**	*2,26*	*1,31*	*603*
		otra	*2,17*	*1,27*	*12*
		Total	*2,26*	*1,31*	*615*
	Total	**catolico**	*2,11*	*1,26*	*1107*
		otra	*1,75*	*,82*	*64*
		Total	*2,09*	*1,24*	*1171*

19.3. Análisis Multivariable de varianza. Cálculo de estadísticos

Contrastes multivariados[b]

Efecto		Valor	F	Gl de la hipótesis	Gl del error gl	Sig.
Intercept	**Traza de Pillai**	,807	2431,309[a]	2,000	1166,000	,000
	Lambda de Wilks	,193	2431,309[a]	2,000	1166,000	,000
	Traza de Hotelling	4,170	2431,309[a]	2,000	1166,000	,000
	Raíz mayor de Roy	4,170	2431,309[a]	2,000	1166,000	,000
RD8A	**Traza de Pillai**	,040	24,341[a]	2,000	1166,000	,000
	Lambda de Wilks	,960	24,341[a]	2,000	1166,000	,000
	Traza de Hotelling	,042	24,341[a]	2,000	1166,000	,000
	Raíz mayor de Roy	,042	24,341[a]	2,000	1166,000	,000
RD8	**Traza de Pillai**	,003	1,600[a]	2,000	1166,000	,202
	Lambda de Wilks	,997	1,600[a]	2,000	1166,000	,202
	Traza de Hotelling	,003	1,600[a]	2,000	1166,000	,202
	Raíz mayor de Roy	,003	1,600[a]	2,000	1166,000	,202
RD8A * RD8	**Traza de Pillai**	,005	3,133[a]	2,000	1166,000	,044
	Lambda de Wilks	,995	3,133[a]	2,000	1166,000	,044
	Traza de Hotelling	,005	3,133[a]	2,000	1166,000	,044
	Raíz mayor de Roy	,005	3,133[a]	2,000	1166,000	,044

a. *Estadístico exacto*

b. *Diseño: Intercept+RD8A+RD8+RD8A * RD8*

19.4. El estadístico F

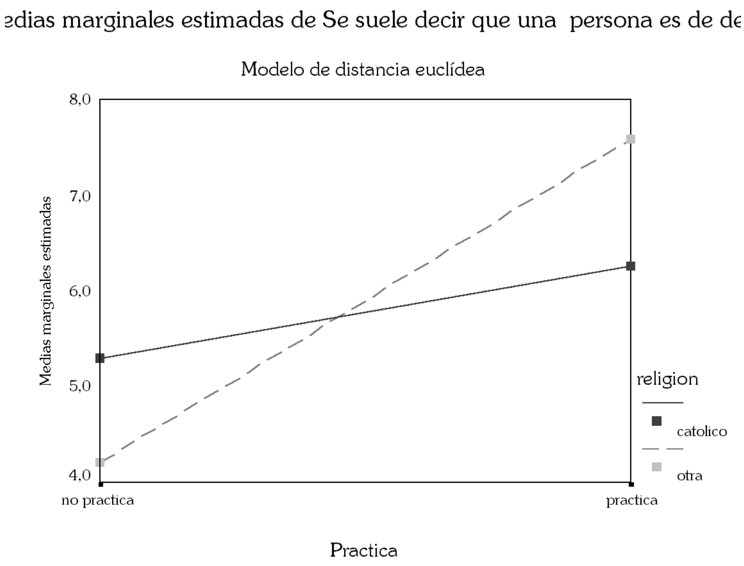
Pruebas de los efectos inter-sujetos

Fuente	Variable dependiente	Suma de cuadrados tipo I	gl	Media cuadrática	F	Sig.
Modelo corregido	**TD9**	41300,563[a]	3	13766,854	11,761	,000
	D3	40,257[b]	3	13,419	8,936	,000
Intercept	**TD9**	2105736,677	1	2105736,677	1799,003	,000
	D3	5109,237	1	5109,237	3402,258	,000
RD8A	**TD9**	32646,033	1	32646,033	27,891	,000
	D3	36,600	1	36,600	24,372	,000
RD8	**TD9**	1402,716	1	1402,716	1,198	,274
	D3	3,343	1	3,343	2,226	,136
RD8A * RD8	**TD9**	7251,814	1	7251,814	6,195	,013
	D3	,314	1	,314	,209	,647
Error	**TD9**	1365975,760	1167	1170,502		
	D3	1752,506	1167	1,502		
Total	**TD9**	3513013,000	1171			
	D3	6902,000	1171			
Total corregido	**TD9**	1407276,323	1170			
	D3	1792,763	1170			

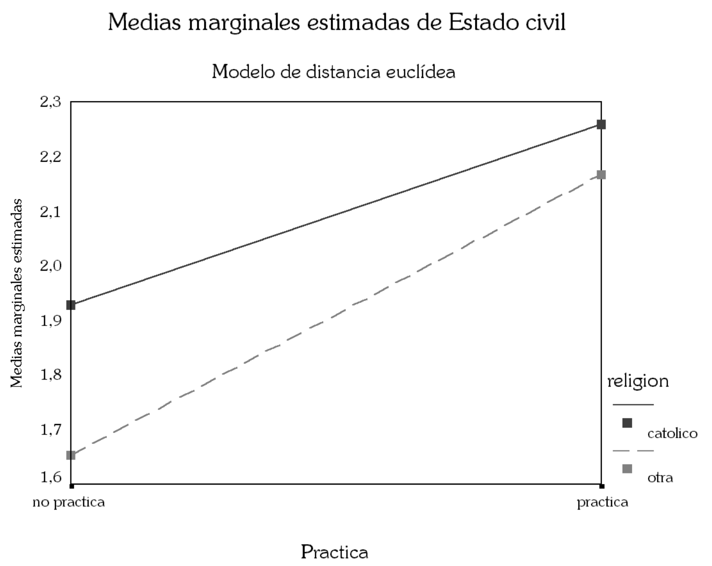
a. *R cuadrado = ,029 (R cuadrado corregido = ,027)*

b. *R cuadrado = ,022 (R cuadrado corregido = ,020)*

19.5. Gráfico de Interacción con la primera variable dependiente



19.6. Gráfico de Interacción con la segunda variable dependiente



19.7. Prueba de LEVENE

Contraste de Levene sobre la igualdad de las varianzas error[a]

	F	gl1	gl2	Sig.
TD9	*5,245*	*8*	*1188*	*,000*
D3	*2,519*	*8*	*1188*	*,010*

Contrasta la hipótesis nula de que la varianza error de la variable dependiente es igual a lo largo de todos los grupos.

a. *Diseño: Intercept+P20+D8+P20 * D8*

20. Bibliografía Comentada

Dentro de esta sección hemos querido presentar dos investigaciones que encadenan una secuencia de modelos estadísticos multivariables entre los que se encuentra el análisis de varianza. Estos trabajos nos recuerdan que las técnicas de investigación social están al servicio de los objetivos de la investigación, de las características de los datos disponibles y de las particularidades de las unidades de observación y que nunca debemos adaptar los segundos a los primeros. Esto es, y parafraseando la máxima maquiavélica, las técnicas de investigación son un medio para conseguir un fin y nuca un fin en sí mismas.

• López, Juan José (1991): "Áreas sociales y población anciana en el municipio de Madrid: aplicación del análisis factorial a un espacio urbano diferenciado", *Economía y sociedad*, nº 5, pp. 79-94.

> *El objetivo de esta investigación es el estudio de la diferenciación espacial de las personas ancianas en Madrid. Partiendo de la hipótesis de que el espacio urbano se encuentra segregado. El método de investigación es el deductivo a partir del enfoque multivariable de la ecología factorial con el que se delimitan las pautas o factores espaciales que sintetizan la multiplicidad de variables que participan en la definición de la calidad de vida del colectivo. Los modelos estadísticos multivariables aplicados han sido: el análisis factorial de componentes principales, factores principales y máxima verosimilitud, a los que se les ha añadido el análisis de correspon-*

dencias; los factores obtenidos se referencian espacialmente aplicando el análisis jerárquico de conglomerados; y, por último, la delimitación ofrecida por el análisis de cluster se verificó aplicando el análisis discriminante. Las variables discriminantes introducidas fueron las puntuaciones factoriales de los factores obtenidos y una variable discreta que recogía el código del tipo del área obtenido. El 85% de los barrios clasificados previamente con el análisis de cluster se encontraban correctamente clasificados. Del total de barrios 21 tuvieron que reclasificarse de nuevo. Por último, se aplica el análisis de varianza con la finalidad de corroborar si las diferencias que se aprecian en el espacio social madrileño de las personas ancianas son aleatorias o evidencian un fenómeno real de diferenciación (H0 se plantea como que no hay diferencias importantes entre los barrios).

- Más, Francisco J. (1999): “Imagen y atracción de centros comerciales suburbanos”, *Investigación y Marketing*, nº 56, pp. 56-65.

La investigación que se presenta tiene como objetivo examinar las percepciones de los consumidores acerca de los centros comerciales y la relación entre las percepciones y la atracción de los mismos. Para el primer objetivo se aplica una doble metodología: (1) análisis descriptivo de las percepciones de los centros comerciales por parte de los consumidores; (2) sobre los valores promedios de las variables en cada centro comercial se aplica un análisis de la varianza (ANOVA) y un análisis múltiple de varianza (MANOVA). Con ello se intenta demostrar la existencia de diferencias significativas entre los distintos centros mediante el examen de su variabilidad: el ANOVA analiza cada una de las variables por separado y el MANOVA todas ellas simultáneamente. Por último, el análisis de la relación entre imagen y atracción de los centros comerciales se realiza aplicando un análisis discriminante al conjunto de variables que describen la imagen de los centros. El objetivo es determinar qué variables diferencian a los individuos que compran asiduamente en los respectivos centros de los que no lo hacen.