

Capítulo 15

Análisis de varianza factorial El procedimiento *Modelo lineal general: Univariante*

Los modelos *factoriales* de análisis de varianza (*factorial = más de un factor*) sirven para evaluar el efecto individual y conjunto de dos o más factores (variables independientes categóricas) sobre una variable dependiente cuantitativa.

Un ANOVA factorial permite estudiar, por ejemplo, si el salario (variable dependiente) de los varones y de las mujeres es diferente (efecto del primer factor) y, al mismo tiempo, si varios grupos de edad tienen distinto salario (efecto del segundo factor). Pero, además, también permite estudiar si las diferencias entre varones y mujeres se repiten o no en cada grupo de edad, es decir, permite determinar si la interacción entre los factores *sexo* y *grupos de edad* afecta a la variable dependiente *salario*. Utilizar más de un factor en un mismo diseño posee la ventaja de poder estudiar el efecto de la interacción entre los factores.

En un modelo de dos factores, los efectos de interés son tres: los dos efectos *principales* (uno por cada factor) y el efecto de la *interacción* entre ambos factores. En un modelo de tres factores, los efectos de interés son siete: los tres efectos principales, los tres efectos de las interacciones dobles (uno por cada interacción entre cada dos factores) y el efecto de la interacción triple (entre los tres factores). Etc.

El procedimiento **Univariante** incluye todos estos modelos factoriales de ANOVA. Pero, además, ofrece la posibilidad de trabajar tanto con factores de efectos fijos como con factores de efectos aleatorios. También permite llevar a cabo *análisis de covarianza* y *análisis de regresión*, y utilizar modelos aleatorizados en bloques y modelos jerárquicos o con factores anidados.

Análisis de varianza factorial

En un análisis de varianza factorial existe una hipótesis nula por cada factor y por cada posible combinación de factores:

- La hipótesis nula referida a un factor afirma que las medias de las poblaciones definidas por los niveles del factor son iguales.
- La hipótesis referida al efecto de una interacción afirma que tal efecto es nulo.

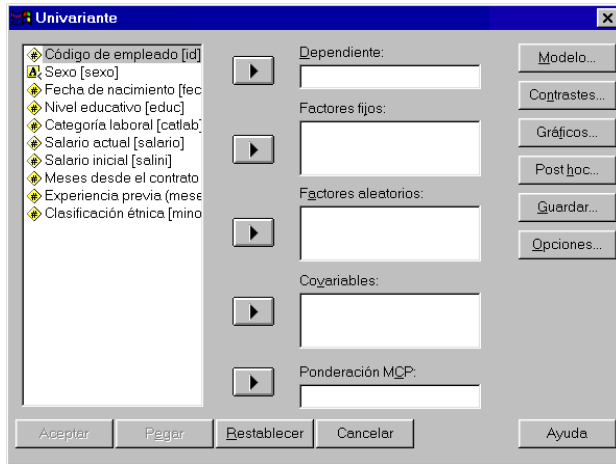
Para contrastar estas hipótesis, el ANOVA factorial se sirve de estadísticos F basados en la lógica ya expuesta en el capítulo anterior al estudiar el modelo de un factor.

Así pues, para cada efecto existe una hipótesis y para cada hipótesis un estadístico F que permite contrastarla. Y al igual que en el ANOVA de un factor, el nivel crítico asociado a cada estadístico F es quien nos permite decidir si podemos mantener o debemos rechazar una hipótesis.

De modo similar a como ocurre en el ANOVA de un factor, en un ANOVA factorial se trabaja con tantas poblaciones como casillas resultan de la combinación de todos los niveles de los factores involucrados. Por ejemplo, en un ANOVA de dos factores, con 2 niveles en un factor y 5 en otro, trabajamos con las $2 \times 5 = 10$ poblaciones definidas por la combinación de niveles de ambos factores. El modelo supone que esas 10 poblaciones son normales (*normalidad*) y que sus varianzas son iguales (homogeneidad de varianzas u homocedasticidad). También supone el modelo que las observaciones han sido aleatoriamente seleccionadas y que, por tanto, son independientes entre sí.

Para llevar a cabo un análisis de varianza de más de un factor:

- ▣ Seleccionar la opción **Modelo lineal general > Univariante...** del menú **Analizar** para acceder al cuadro de diálogo *Univariante* que muestra la figura 15.1.

Figura 15.1. Cuadro de diálogo *Univariante*.

La lista de variables contiene un listado con todas las variables del archivo de datos, incluidas las que poseen formato de cadena. Para obtener un ANOVA factorial con las especificaciones que el procedimiento **Univariante** tiene establecidas por defecto:

- ▣ Seleccionar una variable *cuantitativa* (de intervalo o razón, y, por tanto, con formato numérico) y trasladarla al cuadro **Dependiente**.
- ▣ Seleccionar dos o más variables *categorías* (nominales u ordinales; con formato numérico o de cadena, indistintamente) y trasladarlas a las listas **Factores fijos** o **Factores aleatorios**. En el caso de utilizar como factor una variable con formato de cadena larga, para distinguir entre los valores de la variable sólo se usan los primeros 8 caracteres de cada valor.

Factores fijos. Un factor de *efectos fijos* es aquel cuyos niveles los establece (fija) el investigador (por ejemplo, *cantidad de fármaco*, con niveles establecidos en 0 mg, 50 mg y 250 mg) o vienen dados por la propia naturaleza del factor (por ejemplo, *sexo*, con niveles varones y mujeres). Los niveles concretos que toma un factor de efectos fijos constituyen la población de niveles sobre los que se hace inferencia.

Factores aleatorios. Un factor de *efectos aleatorios* es aquel cuyos niveles son seleccionados de forma aleatoria entre todos los posibles niveles del factor (por ejemplo, *cantidad de fármaco*, con niveles 17 mg, 172 mg y 223 mg, obtenidos seleccionándolos aleatoriamente entre todos los posibles niveles de 0 a 250 mg). Los niveles concretos que toma un factor de efectos aleatorios constituyen sólo una muestra de la población de niveles sobre los que se hace inferencia.

Covariables. En el caso de que se desee llevar a cabo un *análisis de covarianza*, la(s) covariable(s) debe(n) trasladarse a esta lista (ver apartado *Análisis de covarianza*).

Ponderación MCP. El modelo lineal general asume que la varianza de la variable dependiente es la misma en todas las poblaciones objeto de estudio (tantas como casillas definidas por la combinación de niveles de los factores). Cuando éste no es el caso (por ejemplo, cuando las casillas con puntuaciones mayores muestran más variabilidad que las casillas con puntuaciones menores), el método ordinario de mínimos cuadrados deja de proporcionar estimaciones óptimas. Si esas diferencias de variabilidad se pueden pronosticar a partir de alguna variable, el método de mínimos cuadrados ponderados (MCP) permite tener en cuenta esa variable al estimar los parámetros de un modelo lineal, dando así más importancia a las observaciones más precisas (es decir, a aquéllas con menos variabilidad). En ese caso, la variable de ponderación debe trasladarse al cuadro **Ponderación MCP**.

Ejemplo (ANOVA factorial)

Este ejemplo muestra cómo llevar a cabo un ANOVA factorial con las especificaciones mínimas que el procedimiento **Univariante** tiene establecidas por defecto.

Vamos a estudiar, en el archivo *Datos de empleados*, si los grupos definidos por la variable *catlab* (categoría laboral), por un lado, y los grupos definidos por la variable *minoría* (clasificación étnica), por otro, difieren en la variable *salario* (salario actual). Para ello:

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), seleccionar la variable *salario* y trasladarla al cuadro **Dependiente**.
- ▶ Seleccionar las variables *catlab* y *minoría* y trasladarlas a la lista **Factores Fijos**.

Aceptando estas selecciones, el *Visor de resultados* ofrece la información que muestran las tablas 15.1 y 15.2.

Tabla 15.1. Factores inter-sujetos.

	Etiquetas de valor	N
Categoría Laboral	1 Administrativo	363
	2 Seguridad	27
	3 Directivo	84
Clasificación étnica	0 No	370
	1 Sí	104

Figura 15.2. Pruebas de los efectos inter-sujetos.

Variable dependiente: Salario actual

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	90341824425,066 ^a	5	18068364885,013	177,742	,000
Intersección	153679697553,394	1	153679697553,394	1511,773	,000
CATLAB	25962509537,754	2	12981254768,877	127,699	,000
MINORÍA	237964814,398	1	237964814,398	2,341	,127
CATLAB * MINORÍA	788578413,074	2	394289206,537	3,879	,021
Error	47574671011,274	468	101655279,939		
Total	699467436925,000	474			
Total corregido	137916495436,340	473			

a. R cuadrado = ,655 (R cuadrado corregido = ,651)

La primera tabla (15.1) muestra el nombre de las variables independientes (factores), sus niveles, incluidas las etiquetas de los valores, y el número de casos que hay en cada grupo.

La tabla resumen del ANOVA (tabla 15.2) contiene la misma información que la tabla resumen del modelo de un factor: las fuentes de variación, las sumas de cuadrados, los grados de libertad (gl), las medias cuadráticas, los estadísticos F y los niveles críticos ($Sig.$) asociados a cada estadístico F . Pero, ahora, toda esa información está referida no sólo a un factor, sino a los tres efectos presentes en un modelo de dos factores.

La fila *Modelo corregido* se refiere a todos los efectos del modelo tomados juntos (el efecto de los dos factores, el de la interacción y el de la constante o intersección). El nivel crítico asociado al estadístico F ($p = 0,000 < 0,05$) nos está diciendo que el modelo explica una parte significativa de la variación observada en la variable dependiente (*salario*). El valor de R^2 (0,655), que se obtiene dividiendo la suma de cuadrados del *Modelo corregido* entre la suma de cuadrados *Total corregida*, indica que los tres efectos incluidos en el modelo (*catlab*, *minoría* y *catlab*minoría*) están explicando el 65,5 % de la varianza de la variable dependiente *salario*.

La fila *Intersección* informa sobre la constante del modelo. Esta constante forma parte del modelo y es necesaria para obtener las estimaciones de las medias de cada casilla. Además, permite contrastar, en el caso de que esto tenga sentido, la hipótesis de que la media total de la variable dependiente vale cero en la población.

La dos filas siguientes recogen los efectos principales, es decir, los efectos individuales de los dos factores incluidos en el modelo: *catlab* y *minoría*. Los niveles críticos ($Sig.$) indican que, mientras los grupos definidos por la variable *catlab* poseen salarios medios significativamente diferentes ($Sig. = 0,000 < 0,05$), los salarios medios de los grupos definidos por la variable *minoría* no difieren ($Sig. = 0,127 > 0,05$).

La siguiente fila contiene información sobre el efecto de la interacción entre *catlab* y *minoría*. El estadístico F correspondiente a este efecto lleva asociado un nivel crítico de 0,021, lo cual indica que la interacción *catlab-minoría* posee un efecto significativo sobre el *salario*. Sólo con este dato, ya podemos anticipar que las diferencias salariales que se dan entre las distintas categorías laborales no son las mismas en los dos grupos étnicos considerados (más adelante precisaremos el significado del efecto de la interacción a través de una serie de estadísticos y de un gráfico de perfil).

La fila *Error* ofrece información relacionada con la fuente de variación *error* o *residual*. La media cuadrática *error* (que vale 101.655.279,939 y que es el divisor en cada cociente F), es un estimador insesgado de la varianza de las 6 poblaciones estudiadas (que suponemos igual en todas ellas).

La penúltima fila (*Total*) muestra la suma de los cuadrados de las puntuaciones de la variable dependiente (información ésta que, de momento, carece de utilidad); sus grados de libertad son el número total de casos utilizados en el análisis.

Y la última fila (*Total corregido*) recoge la variación total: la variación debida a cada efecto más la variación error.

Comparaciones *post hoc* o *a posteriori*

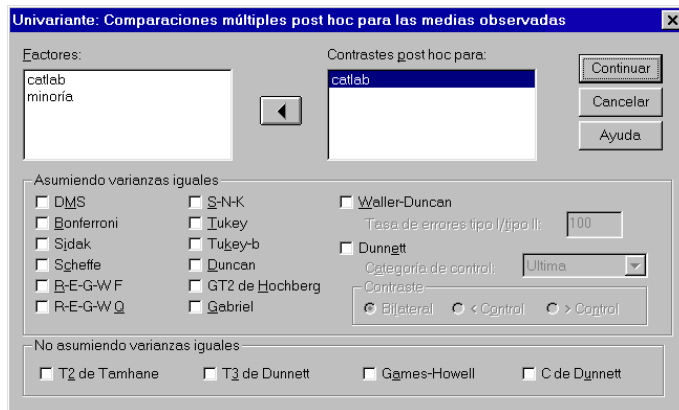
Si alguno de los estadísticos F correspondientes a los efectos principales resulta significativo, puede interesar efectuar comparaciones *post hoc*. Recordemos que los estadísticos F del ANOVA únicamente nos permiten contrastar la hipótesis general de que los promedios comparados son iguales. Al rechazar esa hipótesis, sabemos que existen diferencias, pero no sabemos dónde se encuentran.

Para saber qué media en concreto difiere de qué otra debemos utilizar un tipo particular de contrastes denominados comparaciones múltiples *post hoc* o comparaciones *a posteriori*. Estas comparaciones permiten controlar la *tasa de error* al efectuar varias comparaciones utilizando las mismas medias, es decir, permiten controlar la probabilidad de cometer errores tipo I al tomar varias decisiones (recordemos que los errores tipo I se cometen cuando se decide rechazar una hipótesis nula que en realidad no debería rechazarse).

Para efectuar comparaciones *post hoc*:

- ▶ Pulsar el botón **Post Hoc...** del cuadro de diálogo *Univariante* (ver figura 15.1) para acceder al subcuadro de diálogo *Univariante: Comparaciones múltiples post hoc* que muestra la figura 15.2.

Figura 15.2. Subcuadro de diálogo *Univariante: Comparaciones múltiples post hoc*.



Este subcuadro de diálogo permite seleccionar las variables independientes cuyos niveles interesa comparar y elegir entre una amplia variedad de procedimientos *post hoc*. Estos procedimientos son los mismos que los ya descritos en el capítulo anterior sobre *ANOVA de un factor*, en el apartado *Comparaciones post hoc*.

Ejemplo (ANOVA factorial > Comparaciones post hoc)

Este ejemplo muestra cómo obtener e interpretar las comparaciones *pot hoc* del procedimiento **Univariante**. Aunque no todos los procedimientos disponibles se basan en la misma lógica, como todos ellos se obtienen e interpretan de la misma forma, basta con marcar uno o dos para estudiar los resultados que genera.

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), trasladar la variable *salario* al cuadro **Dependiente** y las variables *catlab* y *minoría* a la lista **Factores fijos**.
- ▶ Pulsar el botón **Post hoc...** para acceder al subcuadro de diálogo *Univariante: Comparaciones múltiples post hoc* (ver figura 15.2).
- ▶ Seleccionar la variable *catlab* en la lista **Factores** y trasladarla a la lista **Contrastes post hoc para**.
- ▶ Marcar la opción **Tukey** del recuadro **Asumiendo varianzas iguales** y la opción **Games-Howell** del recuadro **No asumiendo varianzas iguales**.

Aceptando estas elecciones, el *Visor de resultados* ofrece los resultados que recogen las tablas 15.3. y 15.4.

La tabla 15.3 muestra el resultado obtenido con los dos procedimientos solicitados: Tukey y Games-Howell. Puede observarse que las conclusiones a las que se llega con ambos procedimientos no son idénticas. De acuerdo con el procedimiento de Tukey, el grupo de *administrativos* no difiere del grupo de *agentes de seguridad* ($Sig. = 0,272$), pero estos dos difieren significativamente del grupo de *directivos* ($Sig. = 0,000$ en ambos casos). De acuerdo con el procedimiento de Games-Howell, todos los grupos difieren significativamente ($Sig. = 0,000$ en las tres comparaciones).

Puesto que el procedimiento de Tukey supone varianzas poblacionales iguales y el Games-Howell no, debemos basar nuestra decisión en aquel que se ajuste a las características de los datos (ver más adelante, en este mismo capítulo, el párrafo *Pruebas de homogeneidad* del apartado *Opciones*).

Tabla 15.3. Comparaciones múltiples *post hoc*.

Variable dependiente: Salario actual

	(I) Categoría Laboral	(J) Categoría Laboral	Diferencia entre medias (I-J)	Error tip.	Sig.	Intervalo de confianza al 95%.	
						Límite inferior	Límite superior
DHS de Tukey	Administrativo	Seguridad	-\$3,100.35	2011,232	,272	-\$7,814.07	\$1,613.38
		Directivo	-\$36,139.26	1220,747	,000	-\$39,000.32	-\$33,278.19
	Seguridad	Administrativo	\$3,100.35	2011,232	,272	-\$1,613.38	\$7,814.07
		Directivo	-\$33,038.91	2230,514	,000	-\$38,266.57	-\$27,811.25
	Directivo	Administrativo	\$36,139.26	1220,747	,000	\$33,278.19	\$39,000.32
		Seguridad	\$33,038.91	2230,514	,000	\$27,811.25	\$38,266.57
Games-Howell	Administrativo	Seguridad	-\$3,100.35	2011,232	,000	-\$4,454.82	-\$1,745.88
		Directivo	-\$36,139.26	1220,747	,000	-\$40,977.02	-\$31,301.50
	Seguridad	Administrativo	\$3,100.35	2011,232	,000	\$1,745.88	\$4,454.82
		Directivo	-\$33,038.91	2230,514	,000	-\$37,881.38	-\$28,196.44
	Directivo	Administrativo	\$36,139.26	1220,747	,000	\$31,301.50	\$40,977.02
		Seguridad	\$33,038.91	2230,514	,000	\$28,196.44	\$37,881.38

Basado en medias observadas. El término error es Error.

Tabla 15.4. Subgrupos homogéneos.

	Categoría Laboral	N	Subconjunto	
			1	2
DHS de Tukey ^a	Administrativo	363	\$27,838.54	\$63,977.80
	Seguridad	27	\$30,938.89	
	Directivo	84		
	Sig.		,222	

a. Alfa = ,05.

La tabla 15.4 muestra un resumen (basado en el procedimiento de Tukey) del resultado obtenido con las comparaciones múltiples. En este resumen, los grupos cuyas medias no difieren entre sí están agrupados en el mismo subconjunto y los grupos cuyas medias difieren forman parte de subconjuntos diferentes. En nuestro ejemplo, vemos que existe un primer subconjunto de grupos homogéneos formado por el grupo de *administrativos* y por el de *agentes de seguridad*; y un segundo subconjunto formado por el grupo de *directivos*.

Gráficos de perfil para la interacción

Las comparaciones múltiples *post hoc* suelen proporcionar toda la información necesaria para poder interpretar correctamente un efecto principal significativo. Pero no ocurre lo mismo con los efectos de las interacciones. La interpretación correcta de una interacción suele requerir la ayuda de un gráfico de líneas, también llamado gráfico de perfil.

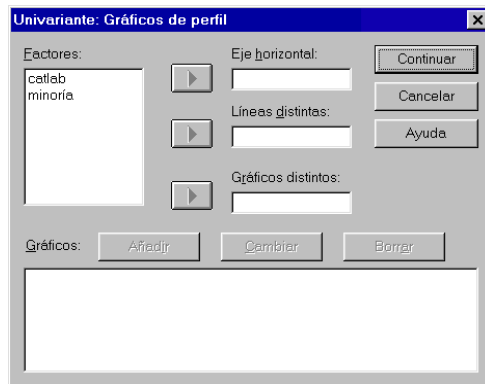
En un gráfico de perfil sobre la interacción entre dos factores, en el eje de ordenadas se representa la escala de las medias de la variable dependiente; en el eje de abscisas se representan los niveles del primer factor; y las líneas del gráfico representan los niveles del segundo factor.

Para representar una interacción triple, es necesario hacer un gráfico de perfil para cada interacción doble en cada nivel del tercer factor. Ir más allá de las interacciones triples no suele tener mucho sentido, entre otras cosas, porque no resulta nada fácil interpretarlas.

Para obtener gráficos de perfil sobre el efecto de las interacciones:

- Pulsar el botón **Gráficos...** del cuadro de diálogo *Univariante* (ver figura 15.1) para acceder al subcuadro de diálogo *Univariante: Gráficos de perfil* que muestra la figura 15.3.

Figura 15.3. Subcuadro de diálogo *Univariante: Gráficos de perfil*.



Este cuadro de diálogo permite obtener gráficos de perfil para combinaciones de dos y tres factores. Para obtener un gráfico de perfil referido a una *interacción doble*:

- ▶ Trasladar a los cuadros **Eje horizontal** y **Líneas distintas** los factores cuya interacción se desea representar.
- ▶ Pulsar el botón **Añadir** para hacer efectiva la selección.
- ▶ Pulsar los botones **Cambiar** y **Borrar** para modificar o eliminar combinaciones previamente añadidas.

Para obtener un gráfico de perfil referido a una *interacción triple*:

- ▶ Antes de pulsar el botón **Añadir**, trasladar el tercer factor al cuadro **Gráficos distintos**.

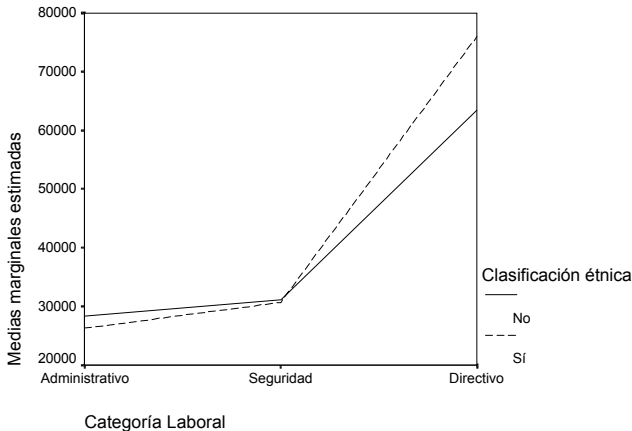
Ejemplo (ANOVA factorial > Gráficos de perfil)

Este ejemplo muestra cómo obtener un gráfico de perfil para el efecto de la interacción entre dos factores.

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), trasladar la variable *salario* al cuadro **Dependiente** y las variables *catlab* y *minoría* a la lista **Factores fijos**.
- ▶ Pulsar el botón **Gráficos...** para acceder al subcuadro de diálogo *Univariante: Gráficos de perfil* (ver figura 15.3).
- ▶ Trasladar la variable *catlab* (categoría laboral) al cuadro **Eje horizontal**.
- ▶ Trasladar la variable *minoría* (clasificación étnica) al cuadro **Líneas distintas**.
- ▶ Pulsar el botón **Añadir** para hacer efectiva la selección de variables.

Aceptando estas elecciones, el *Visor de resultados* construye un gráfico de líneas como el que muestra la figura 15.4.

Figura 15.4. Gráfico de perfil de *catlab* por *minoría*.



En el gráfico de la figura 15.4 aparecen representadas las *medias* de *salario* calculadas en cada subgrupo resultante de combinar cada nivel de la variable *catlab* con cada nivel de la variable

minoría. Una rápida inspección de las líneas nos pone en la pista del significado de esta interacción: en principio, tanto en el grupo de blancos (*minoría = no*) como en el de no blancos (*minoría = sí*), el salario medio parece mayor en los *agentes de seguridad* que en los *administrativos*, y mayor todavía en los *directivos* que en los *agentes de seguridad*; sin embargo, mientras en los grupos de *administrativos* y de *agentes de seguridad* no parecen existir diferencias entre los dos grupos de clasificación étnica, en el grupo de *directivos* las diferencias entre los dos grupos de clasificación étnica parecen evidentes. Por tanto, las diferencias en salario entre los dos grupos de *clasificación étnica* parece no ser la misma a lo largo de todas las *categorías laborales*.

Ese es el significado de la interacción. Pero una interpretación correcta del efecto de la interacción requiere utilizar *comparaciones múltiples* para determinar qué medias en concreto difieren de qué otras. Desafortunadamente, los cuadros de diálogo del procedimiento **Univariante** no contemplan la posibilidad de llevar a cabo estas comparaciones múltiples, por lo que es necesario utilizar la sintaxis SPSS (ver más adelante, en este mismo capítulo, el párrafo *Comparar los efectos principales* del apartado *Opciones*).

Análisis de covarianza

El *análisis de covarianza* (ANCOVA) es una técnica de control estadístico que permite eliminar de la variable dependiente del ANOVA el efecto atribuible a variables no incluidas en el diseño como factores y, por tanto, no sometidas a control experimental.

La forma de controlar estadísticamente el efecto de estas variables *extrañas* consiste en efectuar un análisis de varianza utilizando como variable dependiente, no las puntuaciones originales de la variable dependiente, sino los errores en los pronósticos resultantes de llevar a cabo un *análisis de regresión lineal* con las covariables como variables *independientes* y la propia variable dependiente del ANOVA como variable *dependiente*.

Para llevar a cabo un análisis de covarianza:

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), seleccionar una variable *cuantitativa* (de intervalo o razón) y trasladarla al cuadro **Dependiente**.
- ▶ Seleccionar dos o más variables *categoricas* (nominales u ordinales) y trasladarlas a las listas **Factores fijos** o **Factores aleatorios**.
- ▶ Seleccionar la(s) variable(s) cuyo efecto se desea controlar y trasladarla(s) a la lista **Covariables** (sólo variables con formato numérico).

En un análisis de covarianza, los efectos de interés siguen siendo los referidos a cada factor y a las interacciones entre factores. Estos efectos se interpretan en los términos ya conocidos: un nivel crítico menor que 0,05 delata la presencia de un efecto significativo.

No obstante, el análisis de covarianza también permite evaluar el efecto individual de cada una de las covariables incluidas en el modelo: el procedimiento **Univariante** ofrece, para cada covariable, un estadístico *F* con su correspondiente nivel crítico. Este estadístico *F* permite contrastar la hipótesis nula de que el coeficiente de regresión correspondiente a una covariable vale cero en la población. Como consecuencia del contraste de esa hipótesis, podemos llegar, como en todo contraste, a dos conclusiones distintas: (1) que una covariable no posee efecto significativo, es decir, que no está linealmente relacionada con la variable dependiente; o (2) que una covariable posee efecto significativo, es decir, que está linealmente relacionada con la variable dependiente.

Una covariable que no posee efecto significativo es una covariable que puede ser eliminada del análisis. De hecho, si todas las covariables utilizadas poseen efectos no significativos, debemos esperar que los resultados del ANCOVA sean similares a los del ANOVA, indicando esto que el control ejercido sobre las covariables resulta innecesario.

Si una o más covariables poseen efectos significativos, pueden ocurrir dos cosas: que el resultado del ANOVA y del ANCOVA sea el mismo, o que sea distinto. Por un lado, si el resultado es el mismo, esto significa que, a pesar de que una o más covariables correlacionan con la variable dependiente, y a pesar de que el efecto atribuible a esas covariables ha sido eliminado de la variación de la variable dependiente, el efecto de las variables independientes sobre la dependiente permanece inalterado (lo que significa que la relación entre la(s) covariable(s) y la variable dependiente no afecta a la relación entre los factores y la variable dependiente). Por otro lado, si el resultado del ANOVA y del ANCOVA es distinto, puede ocurrir que lo sea por dos motivos: porque un efecto significativo del ANOVA ha pasado a ser no significativo en el ANCOVA, o porque un efecto no significativo del ANOVA ha pasado a significativo en el ANCOVA. En el primer caso (efecto significativo que deja de serlo), podemos interpretar que el efecto detectado en el ANOVA debe ser atribuido a la(s) covariable(s) incluidas en el análisis y no a la variable independiente; en el segundo caso (efecto no significativo que pasa a serlo), podemos interpretar que la variable independiente en cuestión, aun no estando relacionada con la variable dependiente globalmente considerada, sí correlaciona con la parte de la variable dependiente no explicada o no atribuible a la(s) covariable(s).

Según se desprende del párrafo anterior, la interpretación apropiada de los resultados de un ANCOVA requiere utilizar como punto de referencia los resultados obtenidos en el correspondiente ANOVA.

Ejemplo (ANOVA factorial > Covariables)

Este ejemplo muestra cómo llevar a cabo e interpretar un análisis de covarianza. Vamos a averiguar si las diferencias observadas en *salario* (salario actual) entre los distintos subgrupos definidos por las variables *catlab* (categoría laboral) y *minoría* (minoría) se mantienen al controlar el efecto (al introducir como *covariables*) de las variables *expprev* (experiencia previa) y *tiempemp* (tiempo desde el contrato). Para ello:

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), trasladar la variable *salario* al cuadro **Dependiente** y las variables *catlab* y *minoría* a la lista **Factores fijos**.
- ▶ Trasladar las variables *expprev* y *tiempemp* a la lista **Covariables**.

Aceptando estas elecciones, el *Visor de resultados* ofrece la información que recoge la tabla 15.5.

Tabla 15.5. Tabla resumen del ANOVA (con covariables).

Variable dependiente: Salario actual

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	92046574962,811 ^a	7	13149510708,973	133,588	,000
Intersección	6558230692,616	1	6558230692,616	66,626	,000
TIEMPEMP	1210038517,568	1	1210038517,568	12,293	,000
EXPPREV	492908058,490	1	492908058,490	5,008	,026
CATLAB	27203135908,843	2	13601567954,422	138,181	,000
MINORÍA	299670835,005	1	299670835,005	3,044	,082
CATLAB * MINORÍA	1091353984,818	2	545676992,409	5,544	,004
Error	45869920473,528	466	98433305,737		
Total	699467436925,000	474			
Total corregido	137916495436,340	473			

a. R cuadrado = ,667 (R cuadrado corregido = ,662)

La información referida a las covariables se encuentra en las filas encabezadas *tiempemp* y *expprev*. Ambas covariables poseen estadísticos con niveles críticos (*Sig.*) menores que 0,05, por lo que podemos afirmar que ambas se encuentran linealmente relacionadas con la variable dependiente *salario*.

A pesar de esto, tras controlar el efecto de estas dos variables, vemos que los tres efectos presentes en el modelo (*catlab*, *minoría* y la interacción *catlab*minoría*) mantienen la misma

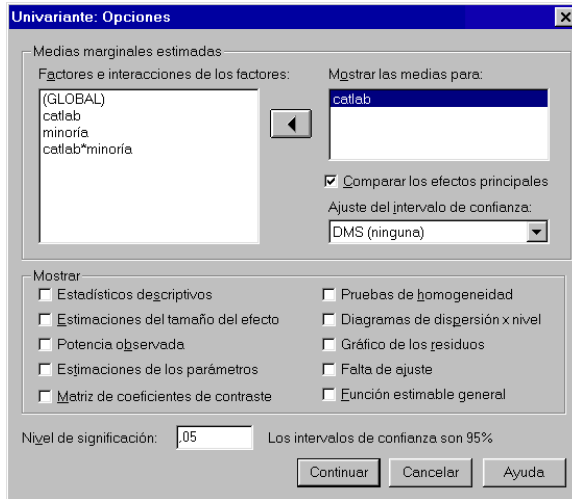
significación que ya tenían en el ANOVA antes de controlar el efecto de las covariables (ver tabla 15.2). Así pues, aunque puede tener sentido controlar el efecto de las variables *tiempemp* y *expprev*, pues ambas están linealmente relacionadas con el variable dependiente *salario*, lo cierto es que los efectos de *catlab*, *minoría* y *catlab*minoría* (que son los efectos que realmente interesa estudiar en un diseño de dos factores) no se alteran por ejercer tal control. Podríamos decir que la relación existente entre las covariables y la variable dependiente no afecta (no altera) la relación existente entre las variables independientes y la dependiente.

Opciones

El subcuadro de diálogo opciones permite obtener información relacionada con varios aspectos complementarios del análisis. Para obtener esta información:

- ▶ Pulsar el botón **Opciones...** (ver figura 15.1) para acceder al subcuadro de diálogo *Univariante: Opciones* que muestra la figura 15.5.

Figura 15.5. Subcuadro de diálogo *Univariante: Opciones*.



Medias marginales estimadas. La lista **Factores e interacción de los factores** muestra un listado con todos los factores incluidos en el diseño y todas las posibles interacciones entre ellos. Trasladando el efecto deseado a la lista **Mostrar las medias para**, el SPSS ofrece una estimación de las medias correspondientes a todos los niveles de ese efecto. Estas medias no son las observadas o empíricas, sino las estimadas a partir de los parámetros del modelo (ver más adelante, dentro de este mismo apartado, la opción *Estimaciones de los parámetros*).

- **Comparar los efectos principales.** Esta opción permite obtener todas las comparaciones dos a dos entre las medias correspondientes a los factores que tienen más de dos niveles y que han sido previamente trasladados a la lista **Mostrar medias para**.

Estas comparaciones por pares se llevan a cabo con la prueba T para dos muestras independientes (ver capítulo 13). Las opciones del menú desplegable **Ajuste del intervalo de confianza** permiten decidir si se desea o no ejercer control sobre la tasa de error (es decir, si se desea o no controlar la probabilidad de cometer errores de tipo I en el conjunto total de comparaciones). La opción por defecto, *DMS (ninguna)*, no ejerce control sobre la tasa de error. La opción *Bonferroni* controla la tasa de error multiplicando el nivel crítico concreto de cada comparación por el número de comparaciones que se están llevando a cabo entre las medias correspondientes a un mismo efecto. La opción *Sidak* corrige la tasa de error mediante $1-(1-p_c)^k$, donde p_c se refiere al nivel crítico de una comparación concreta y k al número de comparaciones.

El procedimiento **Univariante** también permite contrastar los efectos *simples*. Es decir, permite comparar entre sí los niveles de un factor dentro de cada nivel del otro factor, lo cual es especialmente útil para interpretar el efecto de la interacción. Pero, para contrastar estos efectos, es necesario recurrir a la sintaxis SPSS. Para ello:

- ▶ En el cuadro de diálogo *Univariante: Opciones* (ver figura 15.5), seleccionar el efecto que contiene la interacción (en el ejemplo, *catlab*minoría*), junto con algún efecto principal, y trasladarlo a la lista **Mostrar las medias para**.
- ▶ Marcar la opción **Comparar los efectos principales**.
- ▶ Pulsar el botón **Continuar** para volver al cuadro de diálogo *Univariante* (ver figura 15.1).
- ▶ Pulsar el botón **Pegar** para pegar en el *Editor de sintaxis* la sintaxis SPSS correspondiente a las elecciones hechas.
- ▶ Modificar la línea “/EMMEANS = TABLES(minoría*catlab)” añadiéndole lo siguiente: “COMPARE(minoría ADJ(BONFERRONI))”.

Con estas correcciones en el archivo de sintaxis obtendremos, además de las medias que el modelo estima para cada casilla (para cada combinación *catlab*minoría*), una tabla con las comparaciones entre cada nivel de *minoría* (clasificación étnica) dentro de cada nivel de *catlab* (categoría laboral). La tabla 15.5 muestra estos resultados, incluyendo el nivel crítico y el intervalo de confianza asociado a cada comparación.

Tabla 15.5. Efectos simples.

Variable dependiente: Salario actual

Categoría laboral	(I) Clasificación étnica	(J) Clasificación étnica	Diferencia entre medias (I-J)	Error típ.	Sig. ^a	Intervalo de confianza al 95 % para diferencia ^a	
						Límite inferior	Límite superior
Administrativo	No	Sí	2096,834	1239,664	,091	-339,163	4532,831
	Sí	No	-2096,834	1239,664	,091	-4532,831	339,163
Seguridad	No	Sí	497,802	3883,391	,898	-7133,240	8128,844
	Sí	No	-497,802	3883,391	,898	-8128,844	7133,240
Directivo	No	Sí	-12662,687*	5165,705	,015	-22813,535	-2511,840
	Sí	No	12662,687*	5165,705	,015	2511,840	22813,54

Basadas en las medias marginales estimadas.

*. La diferencia de las medias es significativa al nivel ,05.

a. Ajuste para comparaciones múltiples: Bonferroni.

Mostrar. Este recuadro contiene información adicional que puede resultar de interés para el analista:

Estadísticos descriptivos. Media, desviación típica y tamaño de cada nivel y de cada combinación de niveles (ver tabla 15.6). (El Visor no construye esta tabla de descriptivos si se seleccionan más de 18 factores).

Tabla 15.6. Estadísticos descriptivos del procedimiento *Univariante: Opciones*.

Salario actual

Categoría Laboral	Clasificación étnica	Media	Dev. típ.	N
Administrativo	No	\$28,341.09	\$7,994.66	276
	Sí	\$26,244.25	\$5,772.87	87
	Total	\$27,838.54	\$7,567.99	363
Seguridad	No	\$31,178.57	\$1,658.74	14
	Sí	\$30,680.77	\$2,562.92	13
	Total	\$30,938.89	\$2,114.62	27
Directivo	No	\$63,374.81	\$18,164.04	80
	Sí	\$76,037.50	\$17,821.96	4
	Total	\$63,977.80	\$18,244.78	84
Total	No	\$36,023.31	\$18,044.10	370
	Sí	\$28,713.94	\$11,421.64	104
	Total	\$34,419.57	\$17,075.66	474

- Estimaciones del tamaño del efecto.** Estimaciones del grado en que cada factor o combinación de factores está afectando a la variable dependiente. El SPSS ofrece el estadístico *eta cuadrado parcial*, que se obtiene, para un efecto concreto E , de la siguiente manera: $(F_E \times gl_E) / (F_E \times gl_E + gl_{error})$; es decir, dividiendo el producto del estadístico F y de los grados de libertad de ese efecto entre ese mismo producto más los grados de libertad del error. Este estadístico se interpreta como *proporción de varianza explicada*: es una estimación de la proporción de variación de la variable dependiente que está explicada por cada efecto. Si se marca esta opción, el *Visor* muestra las estimaciones del tamaño de los efectos en la tabla resumen del ANOVA (ver tabla 15.7).

Tabla 15.7. Tabla resumen del ANOVA, con las estimaciones del tamaño de los efectos.

Variable dependiente: Salario actual

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta cuadrado	Potencia observada
Modelo corregido	90341824425,066 ^a	5	18068364885,013	177,742	,000	,655	1,000
Intersección	153679697553,394	1	153679697553,394	1511,773	,000	,764	1,000
CATLAB	25962509537,754	2	12981254768,877	127,699	,000	,353	1,000
MINORÍA	237964814,398	1	237964814,398	2,341	,127	,005	,333
CATLAB * MINORÍA	788578413,074	2	394289206,537	3,879	,021	,016	,700
Error	47574671011,274	468	101655279,939				
Total	699467436925,000	474					
Total corregido	137916495436,340	473					

a. Calculado con alfa = ,05

b. R cuadrado = ,655 (R cuadrado corregido = ,651)

- Potencia observada.** Estimaciones de la potencia asociada al contraste de cada efecto. La potencia observada de un contraste se refiere a la capacidad de ese contraste para detectar una diferencia poblacional tan grande como la diferencia muestral de hecho observada. El SPSS calcula el valor de la potencia utilizando un nivel de significación de 0,05, pero este valor puede cambiarse mediante la opción **Nivel de significación** que se encuentra dentro de este mismo cuadro de diálogo. La potencia de cada efecto aparece en la tabla resumen del ANOVA (ver tabla 15.7).
- Estimaciones de los parámetros.** Los modelos de ANOVA contienen una serie de parámetros a partir de los cuales se obtienen las medias que el modelo estima para cada nivel o combinación de niveles. Al marcar esta opción, el SPSS ofrece la información que recoge la tabla 15.8.

Tabla 15.8. Estimaciones de los parámetros del modelo.

Variable dependiente: Salario actual

Parámetro	B	Error típ.	t	Sig.	Intervalo de confianza al 95%.	
					Límite inferior	Límite superior
Intersección	76037,500	5041,212	15,083	,000	66131,287	85943,713
[CATLAB=1]	-49793,247	5155,800	-9,658	,000	-59924,630	-39661,864
[CATLAB=2]	-45356,731	5764,847	-7,868	,000	-56684,919	-34028,543
[CATLAB=3]	0 ^a	,	,	,	,	,
[MINORÍA=0]	-12662,687	5165,705	-2,451	,015	-22813,535	-2511,840
[MINORÍA=1]	0 ^a	,	,	,	,	,
[CATLAB=1] * [MINORÍA=0]	14759,522	5312,370	2,778	,006	4320,470	25198,573
[CATLAB=1] * [MINORÍA=1]	0 ^a	,	,	,	,	,
[CATLAB=2] * [MINORÍA=0]	13160,490	6462,603	2,036	,042	461,178	25859,801
[CATLAB=2] * [MINORÍA=1]	0 ^a	,	,	,	,	,
[CATLAB=3] * [MINORÍA=0]	0 ^a	,	,	,	,	,
[CATLAB=3] * [MINORÍA=1]	0 ^a	,	,	,	,	,

a. Al parámetro se le ha asignado el valor cero porque es redundante.

Las estimaciones de las medias se obtienen combinando los parámetros involucrados en la obtención de cada media. Por ejemplo, la estimación de la media de los *administrativos blancos* se obtiene sumando: el valor de la *Intersección* (76.037,500), el valor correspondiente a *administrativos* ($[CATLAB=1] = -49.793,247$), el valor correspondiente a *blancos* ($[MINORÍA=0] = -12.662,687$) y el valor correspondiente a los *administrativos blancos* ($[CATLAB=1] * [MINORÍA=0] = 14.759,522$). Se obtiene así un salario medio de 28.341,09, que es el valor que muestra para los *administrativos blancos* la tabla 15.7. Puesto que las estimaciones de los parámetros suman cero para cada efecto, la tabla no recoge las estimaciones que resultan redundantes.

La tabla 15.8 también muestra el error típico asociado a cada estimación y un estadístico *t* que permite contrastar la hipótesis de que un determinado parámetro vale cero en la población. Cada estimación, además, aparece acompañada del intervalo de confianza calculado al 95 por ciento.

- Matriz de coeficientes de contraste.** Permite obtener la matriz **L** con los coeficientes asociados a cada efecto (coeficientes que definen el conjunto de hipótesis presentes en un determinado modelo).

- **Pruebas de homogeneidad.** Ofrece el estadístico de Levene sobre homogeneidad de varianzas, el cual permite contrastar la hipótesis de que la varianza de la variable dependiente es la misma en el conjunto de poblaciones definidas por la combinación de factores. En nuestro ejemplo, la tabla 15.9 muestra, para el estadístico de Levene, un valor de 24,72, con un nivel crítico asociado de 0,000. Puesto que el nivel crítico es muy pequeño, debemos rechazar la hipótesis de homogeneidad o igualdad de varianzas.

Tabla 15.9. Estadístico de Levene sobre igualdad de varianzas.

	F	gl1	gl2	Sig.
Salario actual	24.720	5	468	.000

- **Diagramas de dispersión por nivel.** Los diagramas de *dispersión por nivel* proporcionan información gráfica sobre la igualdad de las varianzas. Ayudan a detectar la posible existencia de algún tipo de relación entre el tamaño de las medias y el de las varianzas. Cuando las varianzas son iguales, los puntos del gráfico se encuentran a la misma altura, es decir, alineados horizontalmente. Las figuras 15.10.a y 15.10.b muestran estos gráficos referidos a las variables *salario actual* (dispersión) y *categoría laboral por minoría* (nivel).

Figura 15.10.a. Diagrama de dispersión (desviación típica) por nivel. Variable dependiente *salario*.

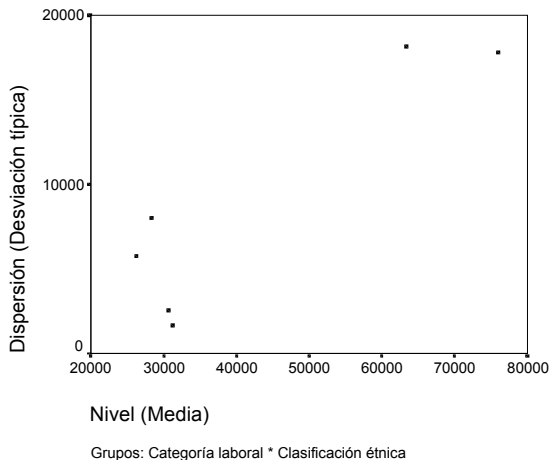
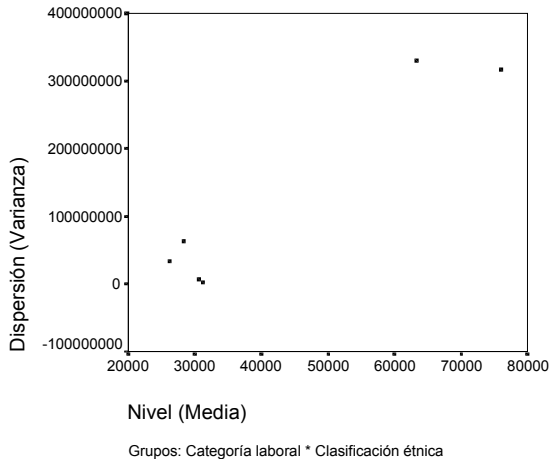


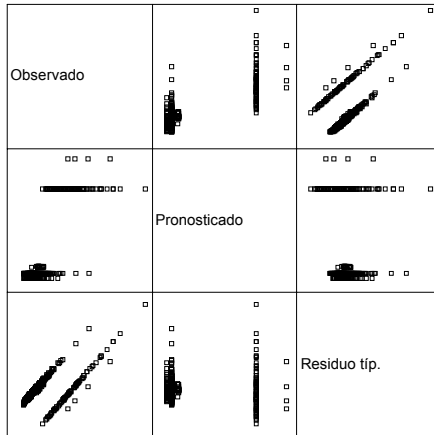
Figura 15.10.b. Diagrama de dispersión (varianza) por nivel. Variable dependiente: *salario*.



Los gráficos muestran 6 puntos, uno por cada nivel resultante de combinar *categoría laboral y minoría*. El hecho de que los puntos no se encuentren horizontalmente alineados indica que las varianzas no son homogéneas (lo cual coincide con la información ofrecida por el estadístico de Levene). Aunque no de forma totalmente clara, parece que las casillas-niveles con medias más grandes son también las casillas-niveles que muestran mayor variación (las dos casillas-niveles con medias más altas muestran mayor variación que el resto de las casillas-niveles).

- **Gráfico de los residuos.** En el contexto del modelo lineal general, los residuos son las diferencias existentes entre los valores observados (las puntuaciones obtenidas en la variable dependiente) y los valores pronosticados por el modelo. En los modelos de ANOVA se supone que los residuos constituyen una variable aleatoria (los residuos son independientes entre sí) y normalmente distribuida. Pero además se supone, según hemos visto ya, homogeneidad de varianzas. El gráfico de los residuos (figura 15.11) nos permite hacernos una idea sobre el cumplimiento de todos estos supuestos, exceptuando el de normalidad.

Figura 15.11. Gráfico de los residuos.



Si los residuos son independientes, el gráfico correspondiente a la relación entre los valores *pronosticados* y los *residuos tipificados* no debe mostrar ninguna pauta de variación sistemática (una línea, una curva, etc.). Y si las varianzas son homogéneas, la dispersión de los *residuos tipificados* debe ser similar a lo largo de todos los valores *pronosticados*. Del gráfico de la figura 15.11 se desprende que, aunque los residuos parecen independientes (pues no muestran una pauta de variación sistemática), la dispersión de los mismos es mayor cuando los valores *pronosticados* son más altos.

Cuando el modelo utilizado ofrece un buen ajuste a los datos, la nube de puntos referida a la relación entre los valores *observados* y los *pronosticados* muestra una pauta de relación claramente lineal. Lógicamente, la pauta es tanto más lineal cuanto mejor ajuste ofrece el modelo.

Nivel de significación. Esta opción permite modificar el nivel de significación con el que El SPSS construye intervalos de confianza y con el que calcula la potencia observada.

Modelos personalizados

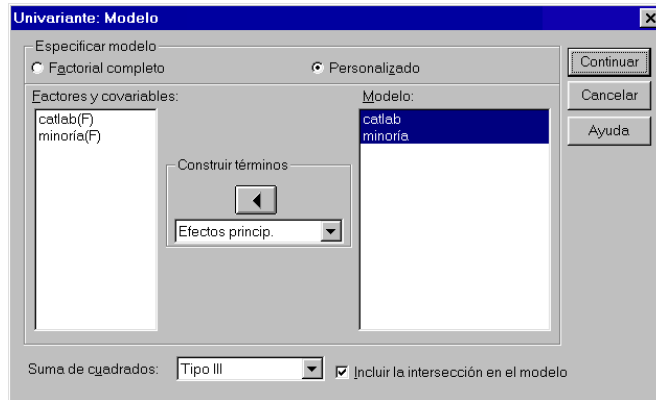
El procedimiento **Univariante** permite ajustar modelos distintos del completamente aleatorizado: modelos sin interacción, modelos con bloques aleatorios, modelos jerárquicos, etc. En este apartado exponemos cómo obtener algunos de estos modelos no completamente aleatorizados.

Especificar un modelo

Para ajustar un modelo personalizado:

- ▶ Pulsar el botón **Modelo...** del cuadro de diálogo *Univariante* (ver figura 15.1) para acceder al subcuadro de diálogo *Univariante: Modelo* que muestra la figura 15.12.

Figura 15.12. Subcuadro de diálogo *Univariante: Modelo*.



- **Factorial completo.** Opción válida para ajustar modelos completamente aleatorizados. Es decir, modelos en los que intervienen todos los efectos principales previamente definidos como factores y todas las posibles interacciones entre ellos. Es la opción que actúa por defecto.

- **Personalizado.** Esta opción permite definir modelos distintos del completamente aleatorizado. Para definir un modelo concreto, debemos seleccionar en la lista **Factores y covariables** los efectos deseados y trasladarlos a la lista **Modelo** utilizando para ello el botón flecha y las opciones del menú desplegable del recuadro **Construir términos**.

Sumas de cuadrados. Este menú desplegable permite elegir entre cuatro métodos distintos de cálculo de las sumas de cuadrados: Tipo I, Tipo II, Tipo III y Tipo IV. Las sumas de cuadrados Tipo III son las más utilizadas y las que el SPSS aplica por defecto:

- En las sumas de cuadrados **Tipo I** (método conocido como *descomposición jerárquica*), cada término se corrige sólo respecto al término que le precede en el modelo.
Se utiliza normalmente en los modelos equilibrados (casillas con tamaños iguales) en los que cualquier efecto principal se evalúa antes que cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se evalúa antes que cualquier efecto de interacción de segundo orden, y así sucesivamente. También se utiliza en los modelos anidados en los que el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente (esta forma de anidado sólo puede definirse mediante sintaxis).
- Las sumas de cuadrados **Tipo II** se obtienen teniendo en cuenta sólo los efectos pertinentes. Un efecto pertinente es un efecto que no está contenido en el efecto que se está evaluando. Para cualesquiera dos efectos E_1 y E_2 , decimos que E_1 está contenido en E_2 si se da alguna de estas tres condiciones: 1) ambos efectos tienen la misma covariable, 2) E_2 consta de más factores que E_1 , y 3) todos los factores presentes en E_1 también lo están en E_2 . Si sólo se evalúan efectos principales (es decir, si el modelo no incluye ningún término de interacción), cada efecto se ajusta teniendo en cuenta el resto de efectos presentes en el modelo (lo que equivale al método *regresión* de obtención de sumas de cuadrados).
Las sumas de cuadrados Tipo II se utilizan normalmente en los modelos equilibrados, en los modelos que sólo contienen efectos principales (no interacciones), y también en los diseños anidados en los que cada efecto especificado está anidado sobre el anterior.
- Las sumas de cuadrados **Tipo III** se calculan ajustando cada efecto teniendo en cuenta cualquier otro efecto que no lo contenga y de forma independiente de cualquier efecto que lo contenga, si existe.
Estas sumas de cuadrados no se alteran por las variaciones del tamaño muestral de las casillas, de modo que son útiles para los modelos no equilibrados sin casillas vacías. Tam-

bién son apropiadas para cualquier modelo que lo sean las sumas de cuadrados tipo I y Tipo II.

- Las sumas de cuadrados **Tipo IV** son apropiadas para analizar tanto modelos equilibrados como no equilibrados cuando existen casillas vacías.

Algunos modelos personalizados

Modelos con bloques aleatorios

Para construir, por ejemplo, un modelo con un factor (*catlab* = categoría laboral) aleatorizado en bloques (definidos por *minoría* = clasificación étnica), debemos tener en cuenta que, en un modelo de este tipo, el factor no interactúa con los bloques, y que, por tanto, los efectos presentes en el modelo son sólo los efectos principales (sin la interacción) de *catlab* y *minoría*. Para definir ese modelo, hay que trasladar a la lista **Modelo** las variables *catlab* y *minoría*, pero no la interacción entre ambas. En general, para definir un modelo con bloques aleatorios (ya sea con un factor o con más de uno):

- En el cuadro de diálogo *Univariante* (ver figura 15.1), seleccionar la variable dependiente y trasladarla al cuadro **Dependiente**.
- Seleccionar tanto las *variables-factores* (las variables independientes del diseño) como las *variables-bloques* (las variables que definen los bloques) y trasladarlas a la lista **Factores fijos**.
- Pulsar el botón **Modelo** para acceder al cuadro de diálogo *Univariante: Modelo* (ver figura 15.12) y marcar la opción **Personalizado**.
- Seleccionar las *variables-factores* y las *variables-bloques* dentro de la lista **Factores y covariables**.
- Seleccionar **Efectos principales** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** las variables previamente seleccionadas.
- Seleccionar ahora, en la lista **Factores y covariables**, sólo las *variables-factores*.
- Seleccionar la opción **Todas de 2** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** todas las posibles combinaciones entre cada dos *variables-factores* (dejando fuera las combinaciones entre las *variables-factores* y las *variables-bloques*, y dejando fuera también las combinaciones de las *variables-bloques* entre sí).

- Si se trata de un modelo con más de dos factores, seleccionar de nuevo, en la lista **Factores y covariables**, sólo las *variables-factores*. Después, seleccionar la opción **Todas de 3** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** todas las posibles combinaciones entre cada tres *variables-factores*. Etc.

Modelos jerárquicos o anidados

En los diseños jerárquicos uno de los factores está anidado en el otro factor. Esto significa que los niveles de uno de los factores son distintos en cada nivel del otro factor. En estas condiciones no es posible evaluar el efecto de la interacción, pero sí los efectos principales. Para ello:

- ▶ Seleccionar las variables independientes y trasladarlas a la lista **Factores fijos**.
- ▶ Pulsar el botón **Modelo** para acceder al cuadro de diálogo *Univariante: Modelo* (ver figura 15.12) y marcar la opción **Personalizado**.
- ▶ Seleccionar la variable que define el factor no anidado dentro de la lista **Factores y covariables**.
- ▶ Seleccionar **Efectos principales** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** esa variable.
- ▶ Pegar las selecciones hechas en una ventana de sintaxis e ir al *Editor de sintaxis* para editar la sintaxis pegada. Si el nombre del factor no anidado es, por ejemplo, *minoría*, la última línea de la sintaxis pegada quedará de esta manera: “DESIGN *MINORÍA*”. Añadir en esa línea el nombre del factor anidado y, a continuación, entre paréntesis, el del factor no anidado. Si el nombre del factor no anidado es, por ejemplo, *catlab*, la línea de sintaxis debe quedar así: “DESIGN *MINORÍA* *CATLAB*(*MINORÍA*)”.

Análisis de regresión lineal

El procedimiento **Univariante** puede utilizarse para ajustar un modelo de regresión lineal y contrastar las hipótesis referidas a los efectos presentes en un modelo de regresión (aunque para esto es más cómodo y apropiado utilizar el procedimiento **Regresión lineal**; ver capítulo 17). Para ajustar un modelo de regresión:

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), seleccionar la variable dependiente y trasladarla al cuadro **Dependiente**.
- ▶ Seleccionar las variables independientes y trasladarlas a la lista **Covariables**.
- ▶ Pulsar el botón **Modelo** para acceder al cuadro de diálogo *Univariante: Modelo* (ver figura 15.12) y seleccionar la opción **Personalizado**.
- ▶ Seleccionar todas las variables de la lista **Factores y covariables**.
- ▶ Seleccionar la opción **Efectos principales** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** las variables previamente seleccionadas.

Con estas especificaciones mínimas obtenemos la tabla resumen del ANOVA con información sobre la significación estadística de cada variable independiente. Para obtener, además, los coeficientes de la ecuación de regresión:

- ▶ En el cuadro de diálogo *Univariante: Opciones* (ver figura 15.5), marcar la opción **Estimaciones de los parámetros**.

Homogeneidad de las pendientes de regresión

En el análisis de covarianza (ANCOVA) ya estudiado en este mismo capítulo, dentro de cada nivel de la variable independiente (o combinación de niveles de las variables independientes), existe una ecuación de regresión referida a la relación entre la variable dependiente y la covariable. Uno de los supuestos del análisis de covarianza es que las pendientes (los coeficientes de regresión) de esas ecuaciones de regresión son homogéneas. Es decir, en el análisis de covarianza se supone que la relación entre la variable dependiente y la covariable es constante a lo largo de todos los grupos definidos por la variable independiente (o por la combinación de niveles de las variables independientes). Para contrastar ese supuesto:

- ▶ En el cuadro de diálogo *Univariante* (ver figura 15.1), seleccionar la variable dependiente y trasladarla al cuadro **Dependiente**.
- ▶ Seleccionar la variable independiente del diseño y trasladarla a la lista **Factores fijos**.
- ▶ Seleccionar la covariable del diseño y trasladarla a la lista **Covariables**.
- ▶ Pulsar el botón **Modelo** para acceder al cuadro de diálogo *Univariante: Modelo* (ver figura 15.12) y marcar la opción **Personalizado**.

- ▶ Seleccionar, dentro de la lista **Factores y covariables**, tanto la variable independiente como la covariable.
- ▶ Seleccionar **Efectos principales** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** las variables previamente seleccionadas.
- ▶ Seleccionar de nuevo la variable independiente y la covariable en la lista **Factores y covariables**.
- ▶ Seleccionar **Interacción** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** el efecto referido a la interacción entre la variable independiente y la covariable.

Procediendo de esta manera es posible contrastar la hipótesis de homogeneidad de las pendientes mediante el estadístico F referido a la interacción (que se interpreta en los términos ya conocidos). Tras esto, si podemos asumir homogeneidad de las pendientes, puede llevarse a cabo un ANCOVA tal como hemos descrito en el apartado *Análisis de covarianza*, en este mismo capítulo, es decir, utilizando un modelo basado en una única estimación de la pendiente de la recta de regresión (lo cual tiene sentido porque suponemos que las pendientes son homogéneas). Pero si rechazamos la hipótesis de homogeneidad de las pendientes, debemos obtener el análisis de covarianza construyendo un modelo personalizado que incorpore estimaciones separadas para cada pendiente de regresión. Para ello:

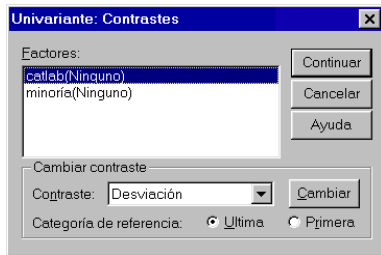
- ▶ En el cuadro de diálogo *Univariante: Modelo* (ver figura 15.12) y marcar la opción **Personalizado**.
- ▶ Seleccionar, dentro de la lista **Factores y covariables**, la variable independiente.
- ▶ Seleccionar **Efectos principales** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** la variable previamente seleccionada.
- ▶ Seleccionar, en la lista **Factores y covariables**, la variable independiente y la covariable.
- ▶ Seleccionar **Interacción** dentro del menú desplegable **Construir términos** y pulsar el botón flecha para trasladar a la lista **Modelo** el efecto referido a la interacción entre la variable independiente y la covariable.
- ▶ Desmarcar la opción **Incluir la intersección en el modelo**.

Contrastes personalizados

Además de las comparaciones *post hoc* (que permiten comparar dos a dos los niveles de un factor) y de las comparaciones referidas a los efectos *simples* (que sirven para interpretar el efecto de la interacción), el procedimiento **Univariante** incluye la posibilidad de efectuar algunos contrastes *a priori*, incluyendo comparaciones de *tendencia*. Para obtener este tipo de contrastes:

- ▶ Pulsar el botón **Contrastes...** del cuadro de diálogo *Univariante* (ver figura 15.1) para acceder al subcuadro de diálogo *Univariante: Contrastes* que muestra la figura 15.13.

Figura 15.13. Subcuadro de diálogo *Univariante: Contrastes*.



La lista **Factores** contiene un listado con los factores previamente seleccionados. Por defecto, los factores no tienen asignado ningún tipo de contraste. Para asignar un tipo de contraste debe utilizarse el menú desplegable **Contraste** y pulsar el botón **Cambiar**. Con todas las opciones disponibles se obtienen $k-1$ comparaciones entre los k niveles de un factor, pero cada contraste permite obtener un tipo particular de comparaciones:

- **Desviación.** Todas las categorías o niveles del factor, excepto uno (el último, por defecto), se comparan con la media total (es decir, con la media de todas las categorías o niveles). En lugar de la última categoría, puede omitirse la primera simplemente seleccionando la opción **Primera** en **Categoría de referencia**. Para omitir una categoría distinta de la primera o la última es necesario utilizar la sintaxis. Por ello, pegar la sintaxis después de marcar las opciones deseadas y añadir en la línea **CONTRASTE**, detrás del nombre del contraste, entre paréntesis, el número de orden (1, 2, 3, etc.) correspondiente a la categoría que se desea omitir (los números de orden de las categorías son 1, 2, 3, etc., aunque los valores

originales de las categorías sean, por ejemplo, 2, 4, 7, etc.). Si se desea omitir, por ejemplo, la segunda categoría del factor *catlab*, la línea CONTRASTE debe quedar de esta manera: “Contrast (catlab) = deviation(2)”.

- **Simple.** Cada categoría, excepto la última, se compara con la última categoría. Seleccionando la opción **Primera en Categoría de referencia**, cada categoría, excepto la primera, se compara con la primera categoría. Para seleccionar una categoría de referencia distinta de la última o la primera es necesario utilizar la sintaxis en la forma descrita en el párrafo anterior.
- **Diferencia.** Cada categoría, excepto la primera, se compara con la media de las categorías anteriores. En los diseños equilibrados, las $k-1$ comparaciones de este contraste son ortogonales.
- **Helmert.** Cada categoría, excepto la última, se compara con la media de las categorías posteriores. En los diseños equilibrados, las $k-1$ comparaciones de este contraste son ortogonales.
- **Repetido.** Cada categoría, excepto la primera, se compara con la categoría anterior.
- **Polinómico.** Comparaciones de tendencia. La primera comparación corresponde a la tendencia lineal; la segunda, a la tendencia cuadrática; etc. En un diseño equilibrado, los contrastes polinómicos son ortogonales. Para más detalles sobre este tipo de comparaciones puede consultarse el párrafo *Polinómico* del apartado *Comparaciones planeadas o a priori*, en el capítulo 14.
- **Especial** (sólo disponible mediante sintaxis). Además de las opciones disponibles en el cuadro de diálogo *Univariante: Contrastes*, también existe la posibilidad de definir cualquier otra comparación entre categorías que pueda resultar de interés. Para ello hay que utilizar la instrucción CONTRAST seguida de la especificación SPECIAL.

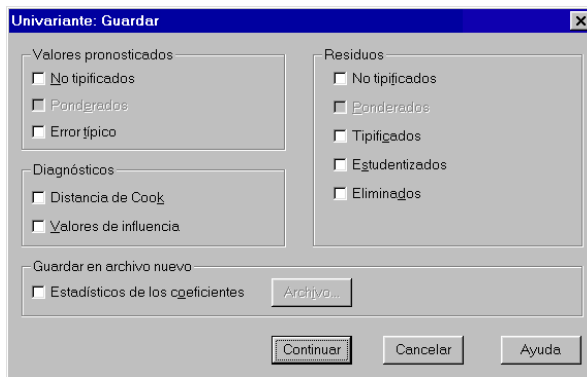
Para comparar, por ejemplo, la primera categoría de *catlab* con la segunda, y la primera con la tercera, debemos utilizar la siguiente línea de sintaxis: “Contrast (catlab) = special (1 -1 0 1 0 -1)”. Para más detalles sobre cómo asignar coeficientes puede consultarse el párrafo *Coefficientes* del apartado *Comparaciones planeadas o a priori*, en el capítulo 14 sobre *ANOVA de un factor*.

Guardar pronósticos y residuos

El procedimiento **Univariante** permite obtener los pronósticos y los residuos que se derivan del modelo que se está intentando ajustar. Para ello:

- ▶ Pulsar el botón **Guardar...** del cuadro de diálogo *Univariante* (ver figura 15.1) para acceder al subcuadro de diálogo *Univariante: Guardar* que muestra la figura 15.14.

Figura 15.14. Subcuadro de diálogo *Univariante: Guardar*.



Todas las opciones de este cuadro de diálogo se explican en el capítulo sobre *Análisis de regresión lineal*, que es donde, quizá, mayor sentido tiene hablar de pronósticos y residuos.