

IBM SPSS Regression 19



Note: Before using this information and the product it supports, read the general information under Notices el p. 47.

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright SPSS Inc. 1989, 2010.**

Prefacio

IBM® SPSS® Statistics es un sistema global para el análisis de datos. El módulo adicional opcional Regresión proporciona las técnicas de análisis adicionales que se describen en este manual. El módulo adicional Regresión se debe utilizar con el sistema básico de SPSS Statistics y está completamente integrado en dicho sistema.

Acerca de SPSS Inc., an IBM Company

SPSS Inc., an IBM Company, es uno de los principales proveedores globales de software y soluciones de análisis predictivo. La gama completa de productos de la empresa (recopilación de datos, análisis estadístico, modelado y distribución) capta las actitudes y opiniones de las personas, predice los resultados de las interacciones futuras con los clientes y, a continuación, actúa basándose en esta información incorporando el análisis en los procesos comerciales. Las soluciones de SPSS Inc. tratan los objetivos comerciales interconectados en toda una organización centrándose en la convergencia del análisis, la arquitectura de TI y los procesos comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de SPSS Inc. como ventaja ante la competencia para atraer, retener y hacer crecer los clientes, reduciendo al mismo tiempo el fraude y mitigando los riesgos. SPSS Inc. fue adquirida por IBM en octubre de 2009. Para obtener más información, visite <http://www.spss.com>.

Asistencia técnica

El servicio de asistencia técnica está a disposición de todos los clientes de mantenimiento. Los clientes podrán ponerse en contacto con este servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de SPSS Inc. o sobre la instalación en alguno de los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia técnica, consulte el sitio web de SPSS Inc. en <http://support.spss.com> o encuentre a su representante local a través del sitio web <http://support.spss.com/default.asp?refpage=contactus.asp>. Tenga a mano su identificación, la de su organización y su contrato de asistencia cuando solicite ayuda.

Servicio de atención al cliente

Si tiene cualquier duda referente a la forma de envío o pago, póngase en contacto con su oficina local, que encontrará en el sitio Web en <http://www.spss.com/worldwide>. Recuerde tener preparado su número de serie para identificarse.

Cursos de preparación

SPSS Inc. ofrece cursos de preparación, tanto públicos como in situ. Todos los cursos incluyen talleres prácticos. Los cursos tendrán lugar periódicamente en las principales ciudades. Si desea obtener más información sobre estos cursos, póngase en contacto con su oficina local que encontrará en el sitio Web en <http://www.spss.com/worldwide>.

Publicaciones adicionales

Los documentos *SPSS Statistics: Guide to Data Analysis*, *SPSS Statistics: Statistical Procedures Companion* y *SPSS Statistics: Advanced Statistical Procedures Companion*, escritos por Marija Norušis y publicados por Prentice Hall, están disponibles y se recomiendan como material adicional. Estas publicaciones cubren los procedimientos estadísticos del módulo SPSS Statistics Base, el módulo Advanced Statistics y el módulo Regression. Tanto si da sus primeros pasos en el análisis de datos como si ya está preparado para las aplicaciones más avanzadas, estos libros le ayudarán a aprovechar al máximo las funciones ofrecidas por IBM® SPSS® Statistics. Si desea información adicional sobre el contenido de la publicación o muestras de capítulos, consulte el sitio web de la autora: <http://www.norusis.com>

Contenido

1	<i>Selección de un procedimiento para la regresión logística binaria</i>	1
2	<i>Regresión Logística</i>	3
	Regresión logística: Establecer regla	5
	Métodos de selección de variables en el análisis de regresión logística	5
	Regresión logística: Definir variables categóricas	6
	Regresión logística: Guardar nuevas variables	7
	Regresión logística: Opciones	9
	Funciones adicionales del comando LOGISTIC REGRESSION	10
3	<i>Regresión logística multinomial</i>	11
	Regresión logística multinomial	13
	Construir términos	14
	Categoría de referencia de regresión logística multinomial	15
	Regresión logística multinomial: Estadísticos	16
	Regresión logística multinomial: Criterios	18
	Opciones de Regresión logística multinomial	19
	Regresión logística multinomial: Guardar	20
	Funciones adicionales del comando NOMREG	21
4	<i>Análisis probit</i>	22
	Análisis probit: Definir rango	24
	Análisis probit: Opciones	24
	Funciones adicionales del comando PROBIT	25

5 Regresión no lineal 26

Lógica condicional (Regresión no lineal) 27
Regresión no lineal: Parámetros 28
Modelos comunes de regresión no lineal 29
Regresión no lineal: Función de pérdida 30
Regresión no lineal: Restricciones para los parámetros 31
Regresión no lineal: Guardar variables nuevas 32
Regresión no lineal: Opciones 32
Regresión no lineal: Interpretación de los resultados 33
Funciones adicionales del comando NLR 34

6 Estimación ponderada 35

Estimación ponderada: Opciones 37
Funciones adicionales del comando WLS 37

7 Regresión por mínimos cuadrados en dos fases 38

Regresión por mínimos cuadrados en dos fases: Opciones 40
Funciones adicionales del comando 2SLS 40

Apéndices

A Esquemas de codificación de variables categóricas 41

Desviación 41
Simple 42
Helmert 43
Diferencia 43
Polinómico 44
Repetido 44
Especial 45
Indicador 46

B Notices

47

Índice

49

Selección de un procedimiento para la regresión logística binaria

Los modelos de regresión logística binaria pueden ajustarse mediante el uso del procedimiento de Regresión logística o del procedimiento de Regresión logística multinomial. Cada uno de estos dos procedimientos contiene opciones que no están disponibles en el otro. Existe entre ambos una distinción teórica importante: el procedimiento de Regresión logística genera todas las predicciones, residuos, estadísticos de influencia y pruebas de bondad de ajuste utilizando los datos a nivel de los casos individuales, independientemente de la forma en que los datos hayan sido introducidos y de si el número de patrones en las covariables es o no menor que el número total de casos; el procedimiento de Regresión logística multinomial, por su parte, agrega los casos de manera interna para formar subpoblaciones con patrones en las covariables idénticos para las variables predictoras, generando predicciones, residuos y pruebas de bondad de ajuste basadas en las citadas subpoblaciones. Si todas las variables predictoras son categóricas, o si alguna variable predictora continua toma sólo un número limitado de valores (de manera que haya varios casos para cada patrón en las covariables), la aproximación mediante subpoblaciones puede generar pruebas de bondad de ajuste válidas y residuos que sean informativos, mientras que el método a nivel de los casos individuales no lo permite.

La opción **Regresión logística** ofrece una serie de funciones únicas que se detallan a continuación:

- Prueba de bondad de ajuste del modelo de Hosmer-Lemeshow
- Análisis por pasos
- Contrastes para definir la parametrización del modelo
- Puntos de corte alternativos para la clasificación
- Gráficos de clasificación
- Aplicación de un modelo ajustado mediante un conjunto de casos sobre otro conjunto de casos reservados
- Almacenamiento de pronósticos, residuos y estadísticos de influencia

La opción **Regresión logística multinomial** ofrece una serie de funciones únicas que se detallan a continuación:

- Pruebas chi-cuadrado de Pearson y de desviación sobre la bondad de ajuste del modelo
- Especificación de subpoblaciones para el agrupamiento de los datos, para las pruebas de bondad de ajuste
- Listado de las frecuencias, frecuencias pronosticadas y residuos por subpoblaciones
- Corrección de las estimaciones de la varianza por sobredispersión

- Matriz de covarianzas para las estimaciones de los parámetros
- Contrastes sobre combinaciones lineales de los parámetros
- Especificación explícita de modelos anidados
- Ajuste de modelos de regresión logística condicional con emparejamiento 1-1 usando variables diferenciadas

Regresión Logística

La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de predictores. Es similar a un modelo de regresión lineal pero está adaptado para modelos en los que la variable dependiente es dicotómica. Los coeficientes de regresión logística pueden utilizarse para estimar la razón de las ventajas (odds ratio) de cada variable independiente del modelo. La regresión logística se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante.

Ejemplo. ¿Qué características del estilo de vida son factores de riesgo de enfermedad cardiovascular? Dada una muestra de pacientes a los que se mide la situación de fumador, dieta, ejercicio, consumo de alcohol, y estado de enfermedad cardiovascular, se puede construir un modelo utilizando las cuatro variables de estilo de vida para predecir la presencia o ausencia de enfermedad cardiovascular en una muestra de pacientes. El modelo puede utilizarse posteriormente para derivar estimaciones de la razón de las ventajas para cada uno de los factores y así indicarle, por ejemplo, cuánto más probable es que los fumadores desarrollen una enfermedad cardiovascular frente a los no fumadores.

Estadísticos. Para cada análisis: Casos totales, Casos seleccionados, Casos válidos. Para cada variable categórica: codificación de los parámetros. Para cada paso: variables introducidas o eliminadas, historial de iteraciones, $-2 \log$ de la verosimilitud, bondad de ajuste, estadístico de bondad de ajuste de Hosmer-Lemeshow, chi-cuadrado del modelo χ^2 , chi-cuadrado de la mejora, tabla de clasificación, correlaciones entre las variables, gráfico de las probabilidades pronosticadas y los grupos observados, chi-cuadrado residual. Para cada variable de la ecuación: coeficiente (B), error típico de B , Estadístico de Wald, razón de las ventajas estimada ($\exp(B)$), intervalo de confianza para $\exp(B)$, log de la verosimilitud si el término se ha eliminado del modelo. Para cada variable que no esté en la ecuación: estadístico de puntuación. Para cada caso: grupo observado, probabilidad pronosticada, grupo pronosticado, residuo, residuo tipificado.

Métodos. Puede estimar modelos utilizando la entrada en bloque de las variables o cualquiera de los siguientes métodos por pasos: Condicional hacia adelante, LR hacia adelante, Wald hacia adelante, Condicional hacia atrás, LR hacia atrás o Wald hacia atrás.

Datos. La variable dependiente debe ser dicotómica. Las variables independientes pueden estar a nivel de intervalo o ser categóricas; si son categóricas, deben ser variables dummy o estar codificadas como indicadores (existe una opción en el procedimiento para recodificar automáticamente las variables categóricas).

Supuestos. La regresión logística no se basa en supuestos distribucionales en el mismo sentido en que lo hace el análisis discriminante. Sin embargo, la solución puede ser más estable si los predictores tienen una distribución normal multivariante. Adicionalmente, al igual que con otras formas de regresión, la multicolinealidad entre los predictores puede llevar a estimaciones sesgadas y a errores típicos inflados. El procedimiento es más eficaz cuando la pertenencia a grupos es una

variable categórica auténtica; si la pertenencia al grupo se basa en valores de una variable continua (por ejemplo “CI alto ” en contraposición a “CI bajo”), deberá considerar el utilizar la regresión lineal para aprovechar la información mucho más rica ofrecida por la propia variable continua.

Procedimientos relacionados. Utilice el procedimiento Diagrama de dispersión para mostrar en pantalla sus datos para multicolinealidad. Si se cumplen los supuestos de normalidad multivariante y de matrices de varianzas-covarianzas iguales, puede obtener una solución más rápida utilizando el procedimiento Análisis discriminante. Si todos los predictores son categóricos, puede además utilizar el procedimiento Loglineal. Si la variable dependiente es continua, utilice el procedimiento Regresión lineal. Puede utilizar el procedimiento Curva COR para realizar gráficos de las probabilidades guardadas con el procedimiento Regresión logística.

Para obtener un análisis de regresión logística

- ▶ Seleccione en los menús:
Analizar > Regresión > Logística binaria...

Figura 2-1
Cuadro de diálogo Regresión logística



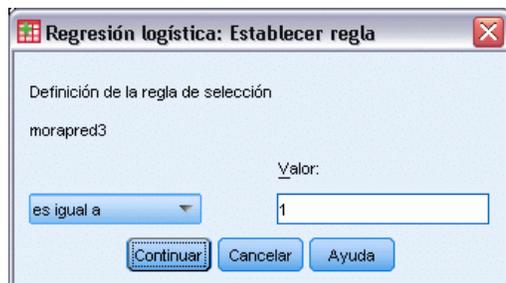
- ▶ Seleccione una variable dependiente dicotómica. Esta variable puede ser numérica o de cadena.
- ▶ Seleccione una o varias covariables. Para incluir términos de interacción, seleccione todas las variables contenidas en la interacción y seleccione >a*b>.

Para introducir variables por grupos (en **bloques**), seleccione las covariables para un bloque y pulse en Siguiente para especificar un nuevo bloque. Repita estos pasos hasta que haya especificado todos los bloques.

Si lo desea, puede seleccionar casos para el análisis. Elija una variable de selección y pulse Regla.

Regresión logística: Establecer regla

Figura 2-2
Cuadro de diálogo Regresión logística: Establecer regla



Los casos definidos por la regla de selección se incluyen en la estimación del modelo. Por ejemplo, si ha seleccionado una variable y la opción igual que y ha especificado 5 como valor, sólo se incluirán en el análisis aquellos casos para los cuales la variable seleccionada tenga un valor igual a 5.

Tanto para los casos seleccionados como para los no seleccionados se generan resultados de clasificaciones y estadísticos. De esta manera, se ofrece un mecanismo para clasificar los nuevos casos basándose en datos ya existentes; o también para realizar la partición de los datos en dos subconjuntos, uno de entrenamiento y otro de prueba, que permiten la validación del modelo generado.

Métodos de selección de variables en el análisis de regresión logística

La selección del método permite especificar cómo se introducen las variables independientes en el análisis. Utilizando distintos métodos se pueden construir diversos modelos de regresión a partir del mismo conjunto de variables.

- **Introducir.** Procedimiento para la selección de variables en el que todas las variables de un bloque se introducen en un solo paso.
- **Selección hacia adelante (Condiciona).** Método de selección por pasos que contrasta la entrada basándose en la significación del estadístico de puntuación y contrasta la eliminación basándose en la probabilidad de un estadístico de la razón de verosimilitud que se basa en estimaciones condicionales de los parámetros.
- **Selección hacia adelante (Razón de verosimilitud).** Método de selección por pasos hacia adelante que contrasta la entrada basándose en la significación del estadístico de puntuación y contrasta la eliminación basándose en la probabilidad del estadístico de la razón de verosimilitud, que se basa en estimaciones de la máxima verosimilitud parcial.
- **Selección hacia adelante (Wald).** Método de selección por pasos hacia adelante que contrasta la entrada basándose en la significación del estadístico de puntuación y contrasta la eliminación basándose en la probabilidad del estadístico de Wald.
- **Eliminación hacia atrás (Condiciona).** Selección hacia atrás por pasos. El contraste para la eliminación se basa en la probabilidad del estadístico de la razón de verosimilitud, el cuál se basa a su vez en las estimaciones condicionales de los parámetros.

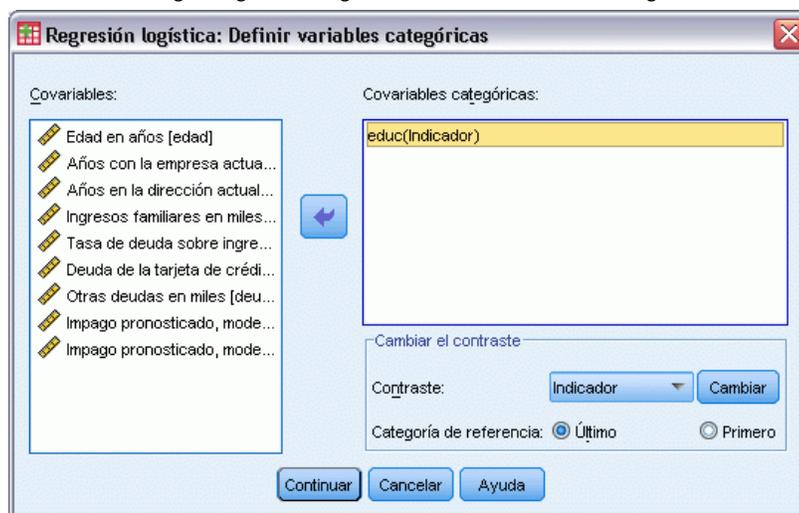
- **Eliminación hacia atrás (Razón de verosimilitud).** Selección hacia atrás por pasos. El contraste para la eliminación se fundamenta en la probabilidad del estadístico de la razón de verosimilitud, el cual se fundamenta en estimaciones de máxima verosimilitud parcial.
- **Eliminación hacia atrás (Wald).** Selección hacia atrás por pasos. El contraste para la eliminación se basa en la probabilidad del estadístico de Wald.

Los valores de significación de los resultados se basan en el ajuste de un único modelo. Por ello, estos valores no suele ser válidos cuando se emplea un método por pasos.

Todas las variables independientes seleccionadas se añaden a un mismo modelo de regresión. Sin embargo, puede especificar distintos métodos de introducción para diferentes subconjuntos de variables. Por ejemplo, puede introducir en el modelo de regresión un bloque de variables que utilice la selección por pasos sucesivos, y un segundo bloque que emplee la selección hacia adelante. Para añadir un segundo bloque de variables al modelo de regresión, pulse en Siguiente.

Regresión logística: Definir variables categóricas

Figura 2-3
Cuadro de diálogo Regresión logística: Definir variables categóricas



Puede especificar los detalles sobre cómo el procedimiento Regresión logística manipulará las variables categóricas:

Covariables. Contiene una lista de todas las covariables especificadas en el cuadro de diálogo principal para cualquier capa, bien por ellas mismas o como parte de una interacción. Si alguna de éstas son variables de cadena o son categóricas, sólo puede utilizarlas como covariables categóricas.

Covariables categóricas. Lista las variables identificadas como categóricas. Cada variable incluye una notación entre paréntesis indicando el esquema de codificación de contraste que va a utilizarse. Las variables de cadena (señaladas con el símbolo < a continuación del nombre) estarán presentes ya en la lista Covariables categóricas. Seleccione cualquier otra covariable categórica de la lista Covariables y muévala a la lista Covariables categóricas.

Cambiar el contraste. Le permite cambiar el método de contraste. Los métodos de contraste disponibles son:

- **Indicador.** Los contrastes indican la presencia o ausencia de la pertenencia a una categoría. La categoría de referencia se representa en la matriz de contraste como una fila de ceros.
- **Simple.** Cada categoría del predictor (excepto la propia categoría de referencia) se compara con la categoría de referencia.
- **Diferencia.** Cada categoría del predictor, excepto la primera categoría, se compara con el efecto promedio de las categorías anteriores. También se conoce como contrastes de Helmert inversos.
- **Helmert.** Cada categoría del predictor, excepto la última categoría, se compara con el efecto promedio de las categorías subsiguientes.
- **Repetidas.** Cada categoría del predictor, excepto la primera categoría, se compara con la categoría que la precede.
- **Polinómico.** Contrastes polinómicos ortogonales. Se supone que las categorías están espaciadas equidistantemente. Los contrastes polinómicos sólo están disponibles para variables numéricas.
- **Desviación.** Cada categoría del predictor, excepto la categoría de referencia, se compara con el efecto global.

Si selecciona Desviación, Simple o Indicador, elija Primera o Última como categoría de referencia. Observe que el método no cambia realmente hasta que se pulsa en Cambiar.

Las covariables de cadena deben ser covariables categóricas. Para eliminar una variable de cadena de la lista Covariables categóricas, debe eliminar de la lista Covariables del cuadro de diálogo principal todos los términos que contengan la variable.

Regresión logística: Guardar nuevas variables

Figura 2-4

Cuadro de diálogo Regresión logística: Guardar nuevas variables



Puede guardar los resultados de la regresión logística como nuevas variables en el conjunto de datos activo:

Valores pronosticados. Guarda los valores pronosticados por el modelo. Las opciones disponibles son Probabilidades y Grupo de pertenencia.

- **Probabilidades.** Para cada caso, guarda la probabilidad pronosticada de aparición del evento. En los resultados, una tabla muestra el nombre y el contenido de cualquier variable nueva.
- **Grupo de pertenencia pronosticado.** Grupo con la mayor probabilidad posterior, basado en puntuaciones discriminantes. El grupo pronosticado por el modelo al cual pertenece el caso.

Influencia. Guarda los valores de estadísticos que miden la influencia de los casos sobre los valores pronosticados. Las opciones disponibles son De Cook, Valores de influencia y DfBeta(s).

- **De Cook.** El análogo, en la regresión logística, al estadístico de influencia de Cook. Una medida de cuánto cambiarían los residuos de todos los casos si un caso particular se excluyera del cálculo de los coeficientes de regresión.
- **Valor de influencia.** La influencia relativa de una observación en el ajuste del modelo.
- **DfBeta(s).** La diferencia en el valor de beta es el cambio en el valor de un coeficiente de regresión que resulta de la exclusión de un caso particular. Se calcula un valor para cada término del modelo, incluyendo la constante.

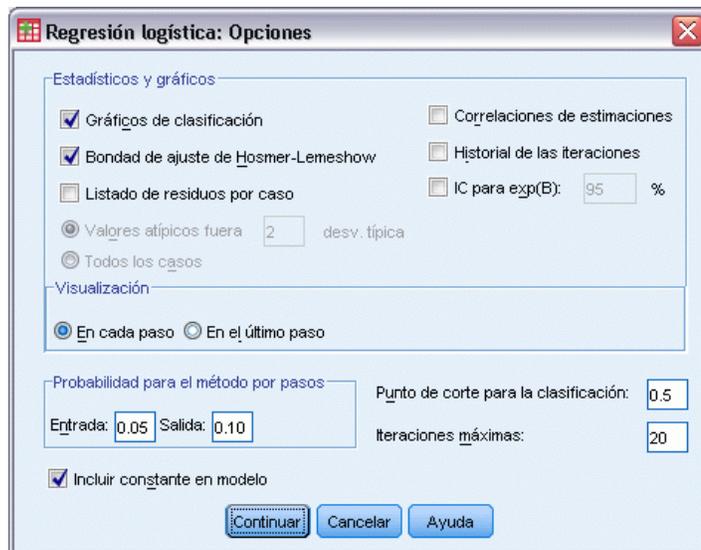
Residuos. Guarda los residuos. Las opciones disponibles son No tipificados, Logit, Método de Student, Tipificados y Desviación.

- **Residuos no tipificados.** Diferencia entre un valor observado y el valor pronosticado por el modelo.
- **Residuo logit.** El residuo del caso si se pronostica en la escala logit. El residuo logit es el residuo dividido por la probabilidad pronosticada multiplicada por 1 menos la probabilidad pronosticada.
- **Residuo estudentizado.** El cambio en la desviación del modelo si se excluye el caso.
- **Residuos tipificados.** El residuo dividido por una estimación de su error típico. Los residuos tipificados, que son conocidos también como los residuos de Pearson o residuos estandarizados, tienen una media de 0 y una desviación típica de 1.
- **Desviación.** Los residuos basados en la desviación del modelo.

Exportar información del modelo a un archivo XML. Las estimaciones de los parámetros y (si lo desea) sus covarianzas se exportan al archivo especificado en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Regresión logística: Opciones

Figura 2-5
Cuadro de diálogo Regresión logística: Opciones



Puede especificar estas opciones para el análisis de regresión logística:

Estadísticos y gráficos. Le permite solicitar estadísticos y gráficos. Las opciones disponibles son Gráficos de clasificación, Bondad de ajuste de Hosmer-Lemeshow, Listado de residuos por caso, Correlaciones de estimaciones, Historial de iteraciones e IC para $\exp(B)$. Seleccione una de las alternativas del grupo Mostrar para mostrar los estadísticos y los gráficos En cada paso o bien sólo para el modelo final, En el último paso.

- **Bondad de ajuste de Hosmer-Lemeshow.** Este estadístico de bondad de ajuste es más robusto que el estadístico de bondad de ajuste tradicionalmente utilizado en la regresión logística, especialmente para los modelos con covariables continuas y los estudios con tamaños de muestra pequeños. Se basa en agrupar los casos en deciles de riesgo y comparar la probabilidad observada con la probabilidad esperada dentro de cada decil.

Probabilidad para el método por pasos. Le permite controlar los criterios por los cuales las variables se introducen y se eliminan de la ecuación. Puede especificar criterios para la Entrada o para la Salida de variables.

- **Probabilidad para el método por pasos.** Una variable se introduce en el modelo si la probabilidad de su estadístico de puntuación es menor que el valor de Entrada, y se elimina si la probabilidad es mayor que el valor de Salida. Para anular los valores por defecto, introduzca valores positivos en los cuadros Entrada y Salida. Entrada debe ser menor que Salida.

Punto de corte para la clasificación. Le permite determinar el punto de corte para la clasificación de los casos. Los casos con valores pronosticados que han sobrepasado el punto de corte para la clasificación se clasifican como positivos, mientras que aquéllos con valores pronosticados menores que el punto de corte se clasifican como negativos. Para cambiar los valores por defecto, introduzca un valor comprendido entre 0,01 y 0,99.

Nº máximo de iteraciones. Le permite cambiar el número máximo de veces que el modelo itera antes de finalizar.

Incluir constante en el modelo. Le permite indicar si el modelo debe incluir un término constante. Si se desactiva, el término constante será igual a 0.

Funciones adicionales del comando LOGISTIC REGRESSION

La sintaxis de comandos también le permite:

- Identificar los resultados por casos mediante los valores o las etiquetas de variable de una variable.
- Controlar el espaciado de los informes de iteración. En lugar de imprimir las estimaciones de los parámetros después de cada iteración, puede solicitar las estimaciones de los parámetros después de cada *n*ésima iteración.
- Cambiar los criterios para finalizar la iteración y para comprobar la redundancia.
- Especificar una lista de variables para los listados por casos.
- Ahorrar memoria manteniendo los datos para cada grupo de segmentación del archivo en un archivo temporal externo durante el procesamiento.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Regresión logística multinomial

La opción Regresión logística multinomial resulta útil en aquellas situaciones en las que desee poder clasificar a los sujetos según los valores de un conjunto de variables predictoras. Este tipo de regresión es similar a la regresión logística, pero más general, ya que la variable dependiente no está restringida a dos categorías.

Ejemplo. Para conseguir una producción y distribución de películas más eficaz, los estudios de cine necesitan predecir qué tipo de películas es más probable que vayan a ver los aficionados. Mediante una regresión logística multinomial, el estudio puede determinar la influencia que la edad, el sexo y las relaciones de pareja de cada persona tiene sobre el tipo de película que prefieren. De esta manera, el estudio puede orientar la campaña publicitaria de una película concreta al grupo de la población que tenga más probabilidades de ir a verla.

Estadísticos. Historial de iteraciones, coeficientes de los parámetros, covarianza asintótica y matrices de correlación, pruebas de la razón de verosimilitud para los efectos del modelo y los parciales, $-2 \log$ de la verosimilitud. Chi-cuadrado de la bondad de ajuste de Pearson y de la desviación. R^2 de Cox y Snell, de Nagelkerke y de McFadden. Clasificación: frecuencias observadas respecto a las frecuencias pronosticadas, por cada categoría de respuesta. Tablas de contingencia: frecuencias observadas y pronosticadas (con los residuos) y proporciones por patrón en las covariables y por categoría de respuesta.

Métodos. Se ajusta un modelo logit multinomial para el modelo factorial completo o para un modelo especificado por el usuario. La estimación de los parámetros se realiza a través de un algoritmo iterativo de máxima verosimilitud.

Datos. La variable dependiente debe ser categórica. Las variables independientes pueden ser factores o covariables. En general, los factores deben ser variables categóricas y las covariables deben ser variables continuas.

Supuestos. Se asume que la razón de ventajas de cualquier par de categorías es independiente de las demás categorías de respuesta. Según esta suposición, por ejemplo, si se introduce un nuevo producto en un mercado, las cuotas de mercado de todos los demás productos se verán afectadas de manera igualmente proporcional. De igual manera, dado un patrón en las covariables, se asume que las respuestas son variables multinomiales independientes.

Obtención de una regresión logística multinomial

- ▶ Seleccione en los menús:
Analizar > Regresión > Logística multinomial ...

Figura 3-1
Cuadro de diálogo Regresión logística multinomial

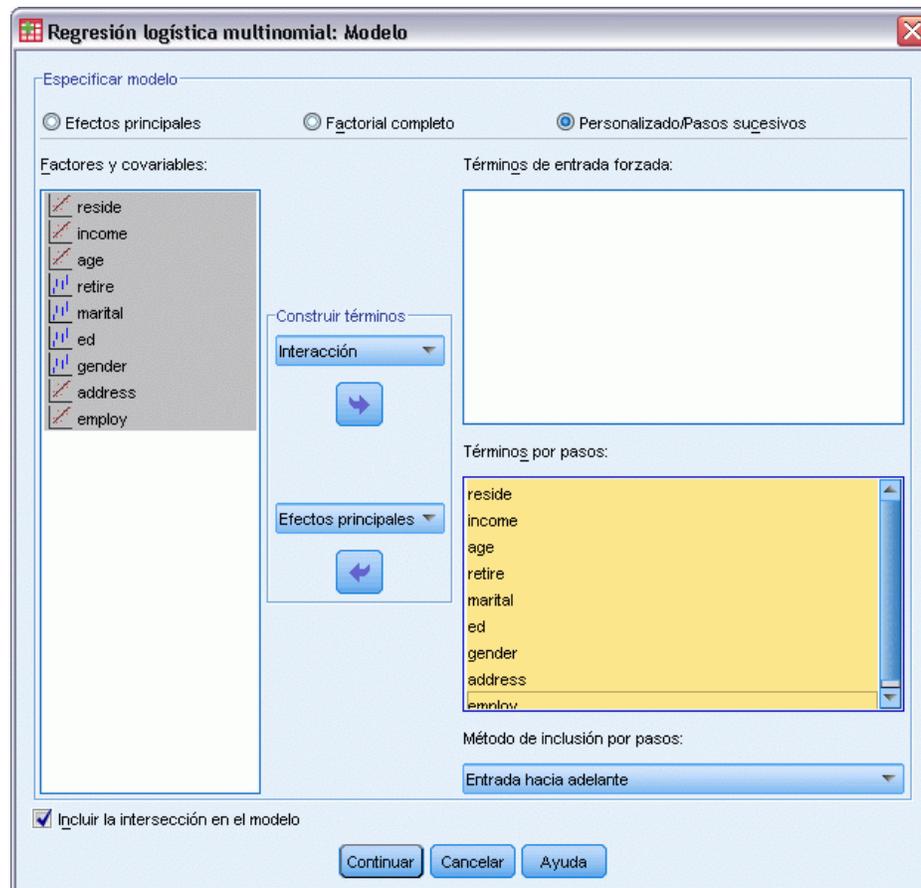


- ▶ Seleccione una variable dependiente.
- ▶ Los factores son opcionales y pueden ser numéricos o categóricos.
- ▶ Las covariables son opcionales, pero si se especifican deben ser numéricas.

Regresión logística multinomial

Figura 3-2

Cuadro de diálogo Regresión logística multinomial: Modelos



Por defecto, el procedimiento de Regresión logística multinomial genera un modelo con los principales efectos que producen las covariables y los factores, pero puede especificar un modelo personalizado o solicitar la selección de un modelo por pasos con este cuadro de diálogo.

Especificar modelo. Un modelo de efectos principales contiene los efectos principales de las covariables y los factores, pero no contiene efectos de interacción. Un modelo factorial completo contiene todos los efectos principales y todas las interacciones factor por factor. No contiene interacciones de covariable. Puede crear un modelo personalizado para especificar subconjuntos de interacciones entre los factores o bien interacciones entre las covariables, o solicitar una selección por pasos de los términos del modelo.

Factores y covariables. Muestra una lista de los factores y las covariables.

Términos de entrada forzada. Los términos añadidos a la lista de entrada forzada siempre se incluyen en el modelo.

Términos por pasos. Los términos añadidos a la lista por pasos se incluyen en el modelo según uno de los métodos por pasos seleccionados por el usuario siguientes:

- **Entrada hacia delante.** Este método se inicia sin términos por pasos en el modelo. En cada paso se añade al modelo el término más significativo, hasta que ninguno de los términos por pasos que quede fuera del modelo tenga una contribución estadísticamente significativa si se añade al modelo.
- **Eliminación hacia atrás.** Este método se inicia introduciendo en el modelo todos los términos especificados en la lista por pasos. En cada paso se elimina del modelo el término menos significativo, hasta que todos los términos por pasos restantes representen una contribución estadísticamente significativa para el modelo.
- **Pasos sucesivos hacia adelante.** Este método se inicia con el modelo que se seleccionaría mediante el método de entrada hacia delante. A partir de ahí, el algoritmo alterna entre la eliminación hacia atrás de los términos por pasos del modelo, y la entrada hacia delante de los términos fuera del modelo. Se sigue así hasta que no queden términos que cumplan con los criterios de entrada o exclusión.
- **Pasos sucesivos hacia atrás.** Este método se inicia con el modelo que se seleccionaría mediante el método de eliminación hacia atrás. A partir de ahí, el algoritmo alterna entre la entrada hacia delante de los términos fuera del modelo, y la eliminación hacia atrás de los términos por pasos del modelo. Se sigue así hasta que no queden términos que cumplan con los criterios de entrada o exclusión.

Incluir la intersección en el modelo. Le permite incluir o excluir del modelo un término de intersección.

Construir términos

Para las covariables y los factores seleccionados:

Interacción. Crea el término de interacción de mayor nivel con todas las variables seleccionadas.

Efectos principales. Crea un término de efectos principales para cada variable seleccionada.

Todas de 2. Crea todas las interacciones dobles posibles de las variables seleccionadas.

Todas de 3. Crea todas las interacciones triples posibles de las variables seleccionadas.

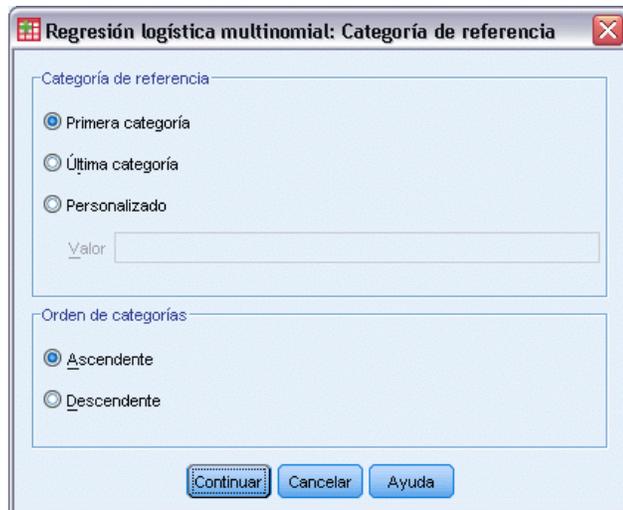
Todas de 4. Crea todas las interacciones cuádruples posibles de las variables seleccionadas.

Todas de 5. Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

Categoría de referencia de regresión logística multinomial

Figura 3-3

Cuadro de diálogo Regresión logística multinomial: Categoría de referencia



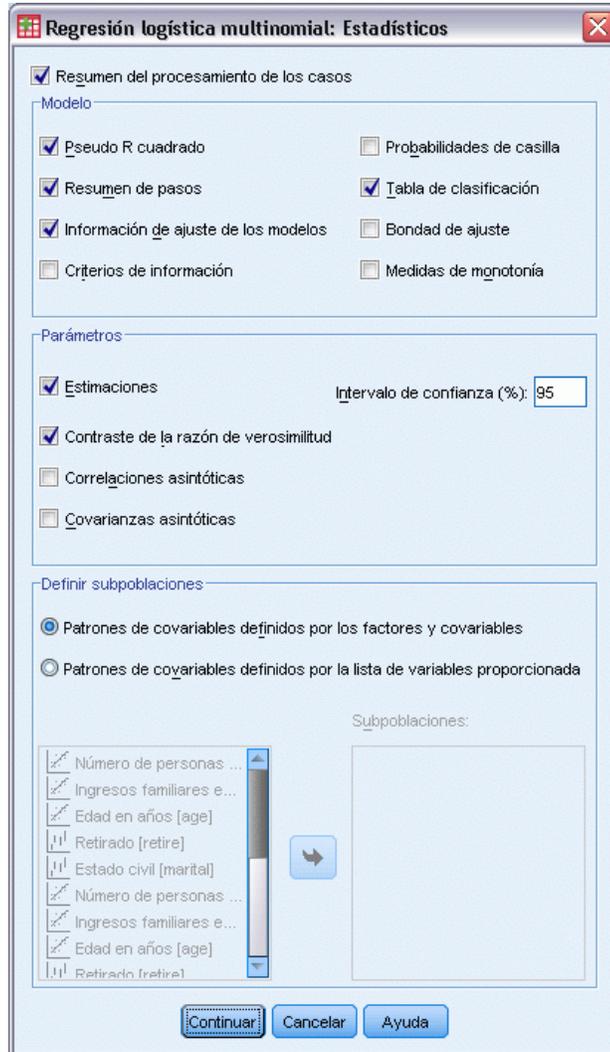
Por defecto, el procedimiento Regresión logística multinomial hace de la última categoría la categoría de referencia. Este cuadro de diálogo le otorga el control sobre la categoría de referencia y sobre la forma de ordenar las categorías.

Categoría de referencia. Especifique la primera, la última o una categoría personalizada.

Orden de categorías. En orden ascendente, el valor mínimo define la primera categoría, y el valor más alto la última. En orden descendente, el valor máximo define la primera categoría y el valor inferior define la última.

Regresión logística multinomial: Estadísticos

Figura 3-4
Cuadro de diálogo Regresión logística multinomial: Estadísticos



Puede especificar los siguientes estadísticos para una regresión logística multinomial:

Resumen de procesamiento de casos. Esta tabla contiene información sobre las variables categóricas especificadas.

Modelo. Estadísticos del modelo global.

- **Pseudo R cuadrado.** Imprime el estadístico de Cox y Snell, de Nagelkerke y el R^2 McFadden.
- **Resumen de pasos.** Esta tabla resume los efectos introducidos o eliminados en cada paso, mediante un método por pasos. No se genera si no se especifica un modelo por pasos en el cuadro de diálogo [Modelo](#).
- **Información de ajuste de los modelos.** Esta tabla compara los modelos ajustado y de sólo intersección o nulo.

- **Criterios de información.** Esta tabla imprime tanto el criterio de información de Akaike (AIC) como el criterio de información bayesiano (BIC).
- **Probabilidades de casilla.** Imprime una tabla de las frecuencias observadas y esperadas (con los residuos) y las proporciones por patrón en las covariables y por categoría de respuesta.
- **Tabla de clasificación.** Imprime una tabla de las respuestas observadas respecto a las respuestas pronosticadas.
- **Estadísticos de bondad de ajuste de chi-cuadrado.** Imprime los estadísticos de chi-cuadrado de Pearson y de chi-cuadrado de la razón de verosimilitud. Los estadísticos se calculan para los patrones en las covariables determinados por todos los factores y las covariables o por un subconjunto de los factores y las covariables definido por el usuario.
- **Medidas de monotonicidad.** Muestra una tabla con información sobre el número de pares concordantes, pares discordantes y empates. La D de Somers, la gamma de Goodman y Kruskal, la tau-a de Kendall y el índice de concordancia C también se muestran en esta tabla.

Parámetros. Estadísticos relativos a los parámetros del modelo.

- **Estimaciones.** Imprime las estimaciones de los parámetros del modelo con un nivel de confianza especificado por el usuario.
- **Contraste de la razón de verosimilitud.** Imprime los contrastes de la razón de verosimilitud para los efectos parciales del modelo. El contraste para el modelo global se imprime de manera automática.
- **Correlaciones asintóticas.** Imprime la matriz de las correlaciones entre las estimaciones de los parámetros.
- **Covarianzas asintóticas.** Imprime la matriz de las covarianzas de las estimaciones de los parámetros.

Definir subpoblaciones. Le permite seleccionar un subconjunto de factores y covariables de manera que pueda definir los patrones en las covariables utilizados por las probabilidades de casilla y las pruebas de bondad de ajuste.

Regresión logística multinomial: Criterios

Figura 3-5

Cuadro de diálogo Regresión logística multinomial: Criterios de convergencia

Regresión logística multinomial: Criterios de convergencia

Iteraciones

Iteraciones máximas: 100

Máxima subdivisión por pasos: 5

Convergencia del logaritmo de la verosimilitud: 0

Convergencia de los parámetros: 0.000001

Imprimir el historial de iteraciones para cada 1 pasos

Comprobar separación de los puntos de los datos desde la iteración 20 hacia adelante

Delta: 0 Tolerancia para la singularidad: 0.00000001

Continuar Cancelar Ayuda

Puede especificar los siguientes criterios para una regresión logística multinomial:

Iteraciones. Le permite especificar el número máximo de veces que desea recorrer el algoritmo, el número máximo de pasos en la subdivisión por pasos, las tolerancias de convergencia para los cambios en el log de la verosimilitud y los parámetros, la frecuencia con que se imprime el progreso del algoritmo iterativo y en qué iteración el procedimiento debe comenzar a comprobar la separación completa o casi completa de los datos.

- **Convergencia del logaritmo de la verosimilitud.** Se asume la convergencia si el cambio absoluto en la función log-verosimilitud es menor que el valor especificado. Este criterio no se aplica si el valor es igual a 0. Especifique un valor no negativo.
- **Convergencia de los parámetros.** Se asume la convergencia si el cambio absoluto en las estimaciones de los parámetros es menor que este valor. Este criterio no se aplica si el valor es igual a 0.

Delta. Le permite especificar un valor no negativo inferior a 1. Este valor se añade a cada casilla vacía de la tabla de contingencia de las categorías de respuesta por patrones de covariables. Se ayuda así a estabilizar el algoritmo y evitar sesgos en las estimaciones.

Tolerancia para la singularidad. Le permite especificar la tolerancia empleada en la comprobación de la singularidad.

Opciones de Regresión logística multinomial

Figura 3-6
Cuadro de diálogo Opciones de Regresión logística multinomial

Puede especificar las siguientes opciones para una regresión logística multinomial:

Escala de dispersión. Le permite especificar el valor de escalamiento de la dispersión que se va a utilizar para corregir la estimación de la matriz de covarianzas de los parámetros. Desviación estima el valor de escalamiento mediante el estadístico de la función de desviación (chi-cuadrado de la razón de verosimilitud). Pearson estima el valor de escalamiento mediante el estadístico chi-cuadrado de Pearson. También puede especificar su propio valor de escalamiento. Debe ser un valor numérico positivo.

Opciones por pasos. Estas opciones le ofrecen el control de los criterios estadísticos cuando se utilizan métodos por pasos para generar un modelo. Se ignoran salvo que se especifique un modelo por pasos en el cuadro de diálogo [Modelo](#).

- **Probabilidad de entrada.** Se trata de la probabilidad del estadístico de la razón de verosimilitud para la entrada de variables. Cuanto mayor sea la probabilidad especificada, más fácil resultará que una variable entre en el modelo. Este criterio se ignora a menos que se seleccione uno de los métodos siguientes: hacia delante, pasos sucesivos hacia adelante o pasos sucesivos hacia atrás.
- **Prueba de entrada.** Éste es el método para introducir los términos en los métodos por pasos. Escoja entre la prueba de la razón de verosimilitud y la prueba de puntuación. Este criterio se ignora a menos que se seleccione uno de los métodos siguientes: hacia delante, pasos sucesivos hacia adelante o pasos sucesivos hacia atrás.

- **Probabilidad de eliminación.** Se trata de la probabilidad del estadístico de la razón de verosimilitud para la eliminación de variables. Cuanto mayor sea la probabilidad especificada, más fácil resultará que una variable permanezca en el modelo. Este criterio se ignora a no ser que se seleccione el método de eliminación hacia atrás, pasos sucesivos hacia adelante o pasos sucesivos hacia atrás.
- **Prueba de eliminación.** Éste es el método utilizado para eliminar términos en los métodos por pasos. Puede elegir entre la prueba de la razón de verosimilitud o la prueba de Wald. Este criterio se ignora a no ser que se seleccione el método de eliminación hacia atrás, pasos sucesivos hacia adelante o pasos sucesivos hacia atrás.
- **Efectos por pasos mínimos en el modelo.** Al utilizar los métodos de eliminación hacia atrás o pasos sucesivos hacia atrás, se especifica el mínimo número de términos que puede incluirse en el modelo. La intersección no se cuenta como término de modelo.
- **Efectos por pasos máximos en el modelo.** Al utilizar los métodos de entrada hacia delante o pasos sucesivos hacia adelante, se especifica el máximo número de términos que puede incluirse en el modelo. La intersección no se cuenta como término de modelo.
- **Restringir jerárquicamente la entrada y la eliminación de términos.** Esta opción permite elegir si se desea aplicar restricciones a la inclusión de términos de modelo. La jerarquía precisa que para que se incluya un término, todos los inferiores que formen parte del que se desea incluir se encuentren antes en el modelo. Por ejemplo, si el requisito de jerarquía está activado, los factores *Estado civil* y *Sexo* deben estar en el modelo antes de poder añadir la interacción *Estado civil*Sexo*. Las tres opciones en forma de botón radial determinan el papel de las covariables a la hora de determinar la jerarquía.

Regresión logística multinomial: Guardar

Figura 3-7
Cuadro de diálogo Regresión logística multinomial: Guardar



El cuadro de diálogo Guardar permite guardar las variables en el archivo de trabajo, así como exportar la información de modelo a un archivo externo.

Variables guardadas.

- **Probabilidades de respuesta estimadas.** Son las probabilidades estimadas de la clasificación de un patrón de factor/covariable en las categorías de respuesta. Hay tantas probabilidades estimadas como categorías de la variable de respuesta; se guardarán hasta 25.
- **Categoría pronosticada.** Es la categoría de respuesta con la mayor probabilidad esperada para un patrón de factor/covariable.
- **Probabilidades de la categoría pronosticada.** Máximo de las probabilidades de respuesta estimadas.
- **Probabilidad de la categoría real.** Es la probabilidad estimada de la clasificación de un patrón de factor/covariable en la categoría observada.

Exportar información del modelo a un archivo XML. Las estimaciones de los parámetros y (si lo desea) sus covarianzas se exportan al archivo especificado en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Funciones adicionales del comando NOMREG

La sintaxis de comandos también le permite:

- Especifique la categoría de referencia de la variable dependiente.
- Incluir los casos con valores perdidos definidos por el usuario.
- Personalizar los contrastes de hipótesis especificando las hipótesis nulas como combinaciones lineales de los parámetros.

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Análisis probit

Este procedimiento mide la relación entre la intensidad de un estímulo y la proporción de casos que presentan una cierta respuesta a dicho estímulo. Es útil para las situaciones en las que se dispone de una respuesta dicotómica que se piensa puede estar influenciada o causada por los niveles de alguna o algunas variables independientes, y es particularmente adecuada para datos experimentales. Este procedimiento le permitirá estimar la intensidad necesaria para que un estímulo llegue a inducir una determinada proporción de respuestas, como la dosis efectiva para la mediana.

Ejemplo. ¿Qué efectividad tiene un nuevo pesticida para matar hormigas y cuál es la concentración adecuada que se debe utilizar? Podría llevar a cabo un experimento en el que se expongan muestras de hormigas a diferentes concentraciones del pesticida y después registrar el número de hormigas muertas y el número de hormigas expuestas. Aplicando el análisis probit a estos datos puede determinar la fuerza de la relación entre concentración y mortalidad, así como determinar la concentración adecuada de pesticida si desea asegurar la exterminación de, por ejemplo, el 95% de las hormigas expuestas.

Estadísticos. Coeficientes de regresión y errores típicos, intersección y su error típico, chi-cuadrado de Pearson de la bondad de ajuste, frecuencias observadas y esperadas e intervalos de confianza para los niveles efectivos de la variable o variables independientes. Gráficos: gráficos de respuestas transformadas.

Este procedimiento utiliza los algoritmos propuestos y aplicados en NPSOL[®] por Gill, Murray, Saunders y Wright para la estimación de los parámetros de modelo.

Datos. Para cada valor de la variable independiente (o para cada combinación de valores para múltiples variables independientes), la variable de respuesta debe contener el recuento del número de casos que presenta la respuesta de interés y que toma dichos valores de la variable independiente, y la variable del total observado debe ser el recuento del número total de casos con dichos valores para la variable independiente. La variable de factor debe ser categórica, codificada como enteros.

Supuestos. Las observaciones deben ser independientes. Si dispone de un gran número de valores diferentes para las variables independientes respecto al número de observaciones, como es probable que suceda en un estudio observacional, puede que no sean válidos los estadísticos de chi-cuadrado y de bondad de ajuste.

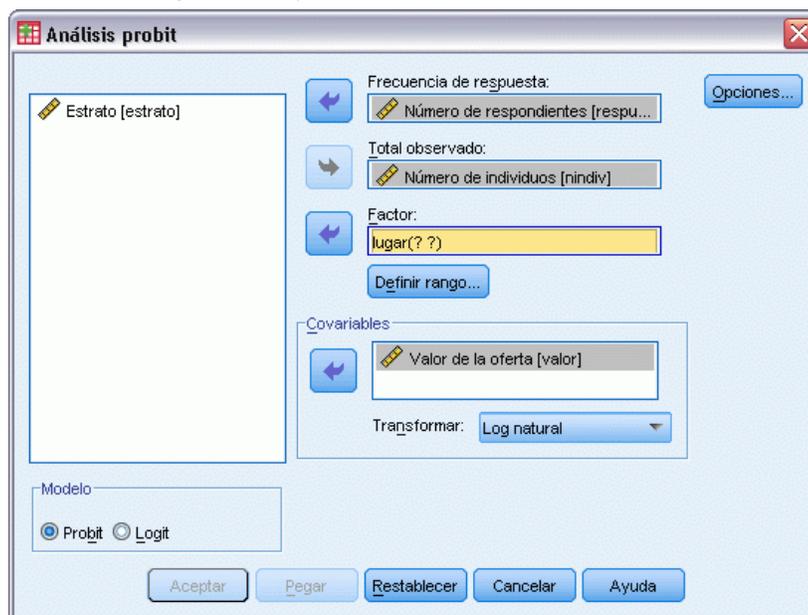
Procedimientos relacionados. El análisis probit está estrechamente relacionado con la regresión logística; de hecho, si elige la transformación logit, este procedimiento calculará esencialmente una regresión logística. En general, el análisis probit es apropiado para los diseños experimentales, mientras que la regresión logística es más adecuada para los estudios observacionales. Las diferencias en los resultados reflejan estas diferencias de énfasis. El procedimiento Análisis

probit informa de las estimaciones de los valores efectivos para las diferentes tasas de respuesta (incluyendo la dosis efectiva para la mediana), mientras que la Regresión logística informa de las estimaciones de las razones de las ventajas (odds ratios) para las variables independientes.

Para obtener un análisis probit

- ▶ En los menús, seleccione:
Analizar > Regresión > Probit...

Figura 4-1
Cuadro de diálogo Análisis probit



- ▶ Seleccione una variable para la frecuencia de respuesta. Esta variable indica el número de casos que presentan una respuesta al estímulo de prueba. Los valores de esta variable no pueden ser negativos.
- ▶ Seleccione una variable para el total observado. Esta variable indica el número de casos a los que se aplicó el estímulo. Para cada caso, los valores de esta variable no pueden ser negativos ni menores que los valores de la variable de frecuencia de respuesta.

Si se desea, puede seleccionarse una variable de factor. Si lo hace, pulse en Definir rango para definir los grupos.

- ▶ Seleccione una o varias covariables. La covariable contiene el nivel del estímulo aplicado en cada observación. Si desea transformar la covariable, seleccione una transformación de la lista desplegable Transformar. Si no se aplica ninguna transformación y hay un grupo de control, éste se incluirá en el análisis.
- ▶ Seleccione el modelo Probit o Logit.

- **Modelo Probit.** Aplica la transformación Probit (la inversa de la función acumulada de la distribución normal típica) a las proporciones de respuesta.
- **Modelo logit.** Aplica la transformación logit (log de las ventajas) a las proporciones de respuesta.

Análisis probit: Definir rango

Figura 4-2
Cuadro de diálogo Análisis probit: Definir rango

Este cuadro de diálogo permite definir el rango de niveles de la variable de factor que serán analizados. Incluye campos de entrada para el mínimo y el máximo, y botones para continuar, cancelar y ayuda.

Mínimo:	<input type="text" value="1"/>
Máximo:	<input type="text" value="3"/>
<input type="button" value="Continuar"/> <input type="button" value="Cancelar"/> <input type="button" value="Ayuda"/>	

Permite especificar los niveles de la variable de factor que serán analizados. Los niveles de factor deben codificarse como enteros consecutivos; se analizarán todos los niveles del rango que especifique.

Análisis probit: Opciones

Figura 4-3
Cuadro de diálogo Análisis probit: Opciones

Este cuadro de diálogo permite configurar las opciones de análisis. Incluye secciones para estadísticos, tasa de respuesta natural y criterios de convergencia.

Estadísticos	
<input checked="" type="checkbox"/>	Frecuencias
<input checked="" type="checkbox"/>	Potencia relativa de la mediana
<input checked="" type="checkbox"/>	Prueba de paralelismo
<input checked="" type="checkbox"/>	Intervalos de confianza fiduciaria
Nivel de significación para el uso del factor de heterogeneidad: <input type="text" value="0.15"/>	
Tasa de respuesta natural	
<input type="radio"/>	Ninguna
<input checked="" type="radio"/>	Calcular a partir de los datos
<input type="radio"/>	Valor: <input type="text"/>
Criterios	
Iteraciones máximas:	<input type="text" value="100"/>
Límite para los pasos:	<input type="text" value="0.1"/>
Tolerancia de la optimalidad:	<input type="text" value="Por defecto"/>
<input type="button" value="Continuar"/> <input type="button" value="Cancelar"/> <input type="button" value="Ayuda"/>	

Puede especificar opciones para el análisis probit:

Estadísticos. Permite solicitar los siguientes estadísticos opcionales: Frecuencias, Potencia relativa de la mediana, Prueba de paralelismo e Intervalos de confianza fiduciaria.

- **Potencia relativa de la mediana.** Muestra la razón de las potencias de las medianas para cada pareja de los niveles del factor. También muestra los límites de confianza al 95% para cada potencia relativa de la mediana. Las potencias relativas de la mediana no están disponibles si no dispone de una variable de factor, o si dispone de más de una covariable.
- **Prueba de paralelismo.** Contraste sobre la hipótesis de que todos los niveles del factor tienen una pendiente común.
- **Intervalos de confianza fiduciaria.** Intervalos de confianza para la dosis del agente requerida para producir una cierta probabilidad de respuesta.

Intervalos de confianza fiduciaria y Potencia relativa de la mediana no están disponibles si se ha seleccionado más de una covariable. Potencia relativa de la mediana y Prueba de paralelismo sólo están disponibles si se ha seleccionado una variable de factor.

Tasa de respuesta natural. Permite indicar una tasa de respuesta natural incluso en la ausencia del estímulo. Ninguna, Calcular a partir de los datos o Valor.

- **Calcular a partir de los datos.** Estima la tasa de respuesta natural a partir de los datos de la muestra. Los datos deben contener un caso que represente el nivel de control, para el cual el valor de las covariables sea 0. Probit estima la tasa de respuesta natural utilizando como valor inicial la proporción de respuestas para el nivel de control.
- **Valor.** Establece la tasa de respuesta natural del modelo (seleccione este elemento cuando conozca de antemano la tasa de respuesta natural). Introduzca la proporción de respuesta natural (la proporción debe ser menor que 1). Por ejemplo, si la respuesta ocurre el 10% de las veces cuando el estímulo es 0, introduzca 0,10.

Criterios. Permite controlar los parámetros del algoritmo iterativo de estimación de los parámetros. Puede anular las opciones por defecto para N° máximo de iteraciones, Límite para los pasos y Tolerancia de la optimalidad.

Funciones adicionales del comando PROBIT

Con el lenguaje de sintaxis de comandos también podrá:

- Solicitar simultáneamente un análisis con ambos modelos, probit y logit.
- Controlar el tratamiento de los valores perdidos.
- Transformar las covariables por bases diferentes de la base 10 o el logaritmo natural.

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Regresión no lineal

Regresión no lineal es un método para encontrar un modelo no lineal para la relación entre la variable dependiente y un conjunto de variables independientes. A diferencia de la regresión lineal tradicional, que está restringida a la estimación de modelos lineales, la regresión no lineal puede estimar modelos con relaciones arbitrarias entre las variables independientes y las dependientes. Esto se lleva a cabo usando algoritmos de estimación iterativos. Tenga en cuenta que este procedimiento no es necesario para los modelos polinómicos simples de la forma $Y = A + BX^{**2}$. Definiendo $W = X^{**2}$, obtenemos un modelo lineal simple, $Y = A + BW$, que se puede estimar usando métodos tradicionales como el procedimiento Regresión lineal.

Ejemplo. ¿Puede pronosticarse la población basándose en el tiempo? Un diagrama de dispersión muestra que parece haber una estrecha relación entre la población y el tiempo, pero la relación es no lineal y por eso exige la utilización de los métodos de estimación especiales del procedimiento Regresión no lineal. Creando una ecuación adecuada, como la del modelo logístico de crecimiento poblacional, podemos obtener una buena estimación del modelo, lo que nos permitirá hacer predicciones sobre la población para épocas que no se han sido medidas.

Estadísticos. Para cada iteración: estimaciones de los parámetros y suma de cuadrados residual. Para cada modelo: suma de cuadrados para regresión, residual, total corregido y no corregido, estimaciones de los parámetros, errores típicos asintóticos y matriz de correlaciones asintóticas de estimaciones de los parámetros.

Nota: La regresión no lineal restringida utiliza los algoritmos propuestos y aplicados en NPSOL[®] por Gill, Murray, Saunders y Wright para la estimación de los parámetros de modelo.

Datos. Las variables dependiente e independientes deben ser cuantitativas. Las variables categóricas, como la religión, la mayoría de edad o el lugar de residencia, han de recodificarse como variables binarias (dummy) o como otro de los tipos de variables de contraste.

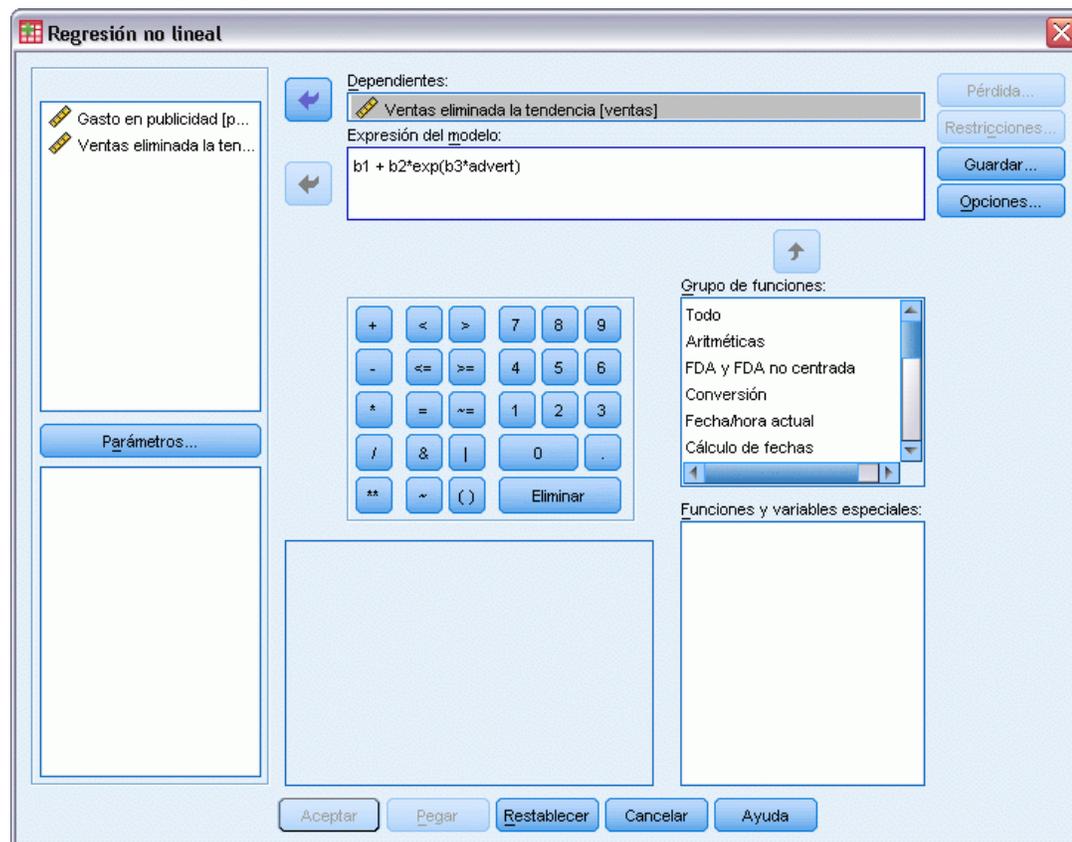
Supuestos. Los resultados son válidos sólo si se ha especificado una función que describa con precisión la relación entre las variables independientes y las dependientes. Además, la elección de buenos valores iniciales es muy importante. Incluso si se ha especificado la forma funcional correcta para el modelo, si no utiliza valores iniciales adecuados, puede que su modelo no logre converger o puede que obtenga una solución que sea óptima localmente en vez de una que sea óptima globalmente.

Procedimientos relacionados. Muchos modelos que en un principio parecen ser no lineales pueden ser transformados en un modelo lineal, el cual pueda ser analizado usando el procedimiento Regresión lineal. Si no está seguro de cuál es el modelo adecuado, el procedimiento Estimación curvilínea puede ayudarle a identificar relaciones funcionales útiles que estén presentes en los datos.

Para obtener un análisis de regresión no lineal

- ▶ En los menús, seleccione:
Analizar > Regresión > No lineal...

Figura 5-1
Cuadro de diálogo Regresión no lineal



- ▶ Seleccione una variable numérica dependiente de la lista de variables del conjunto de datos activo.
- ▶ Para construir una expresión para el modelo, introduzca la expresión en el campo Expresión del modelo o bien pegue en el campo los componentes (variables, parámetros, funciones).
- ▶ Identifique los parámetros del modelo pulsando en Parámetros.

Un modelo segmentado (uno que adquiere diferentes formas en distintas partes de su dominio) se debe especificar usando la lógica condicional dentro de la declaración única del modelo.

Lógica condicional (Regresión no lineal)

Se puede especificar un modelo segmentado utilizando la lógica condicional. Para usar la lógica condicional dentro de la expresión del modelo o de la función de pérdida, debe construir la suma de una serie de términos, uno para cada condición. Cada término se compone de una expresión

lógica (entre paréntesis) multiplicada por la expresión que resultará cuando esa expresión lógica es verdadera.

Por ejemplo, considere un modelo segmentado que sea igual a 0 para $X \leq 0$, X para $0 < X < 1$ y 1 para $X \geq 1$. La expresión para este ejemplo es:

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

Todas las expresiones lógicas entre paréntesis deben ser evaluables como 1 (verdadero) o 0 (falso). Así:

Si $X \leq 0$, la anterior se reduce a $1 * 0 + 0 * X + 0 * 1 = 0$.

Si $0 < X < 1$, se reduce a $0 * 0 + 1 * X + 0 * 1 = X$.

Si $X \geq 1$, se reduce a $0 * 0 + 0 * X + 1 * 1 = 1$.

Se pueden construir con facilidad ejemplos más complicados reemplazando diferentes expresiones lógicas y expresiones de resultado. Recuerde que las desigualdades dobles, como $0 < X < 1$, deben escribirse como expresiones compuestas, de la forma $(X > 0 \ \& \ X < 1)$.

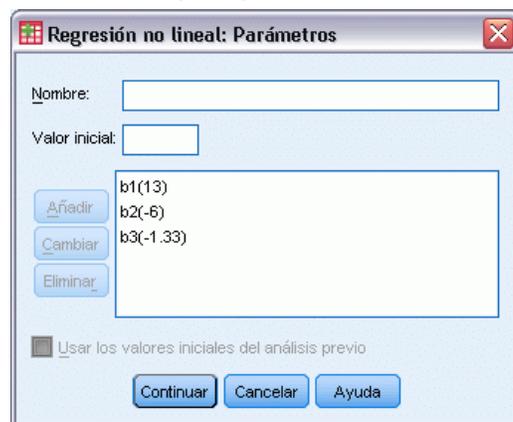
Se pueden utilizar variables de cadena dentro de las expresiones lógicas:

$$(\text{ciudad} = \text{'Madrid'}) * \text{costliv} + (\text{ciudad} = \text{'Guadalajara'}) * 0.59 * \text{costliv}$$

Esto da lugar a una expresión (el valor de la variable *costliv*) para los madrileños y a otra (el 59% de ese valor) para los habitantes de Guadalajara. Las constantes de cadena deben ir entre comillas o apóstrofes, como se muestra aquí.

Regresión no lineal: Parámetros

Figura 5-2
Cuadro de diálogo Regresión no lineal: Parámetros



Los parámetros son las partes del modelo que son estimadas por el procedimiento Regresión no lineal. Los parámetros pueden ser constantes aditivas, coeficientes multiplicativos, exponentes o valores usados para evaluar funciones. Todos los parámetros que hayan sido definidos aparecerán (con sus valores de inicio) en la lista Parámetros del cuadro de diálogo principal.

Nombre. Debe especificarse un nombre para cada parámetro. Debe ser un nombre de variable válido y debe ser el nombre utilizado en la expresión del modelo del cuadro de diálogo principal.

Valor inicial. Permite especificar un valor de inicio para el parámetro, preferiblemente lo más próximo posible a la solución final esperada. Los valores iniciales no adecuados pueden dar como resultado un fallo de convergencia o una convergencia sobre una solución local (en vez de global) o físicamente imposible.

Usar los valores iniciales del análisis previo. Si ya se ha ejecutado una regresión no lineal desde este cuadro de diálogo, puede seleccionar esta opción para obtener los valores iniciales de los parámetros a partir de sus valores en la ejecución previa. De esta forma podrá continuar buscando cuando el algoritmo está convergiendo lentamente. Los primeros valores iniciales seguirán apareciendo en la lista Parámetros del cuadro de diálogo principal.

Nota: Esta selección persistirá en este cuadro de diálogo durante el resto de la sesión. Si cambia el modelo, asegúrese de desactivarla.

Modelos comunes de regresión no lineal

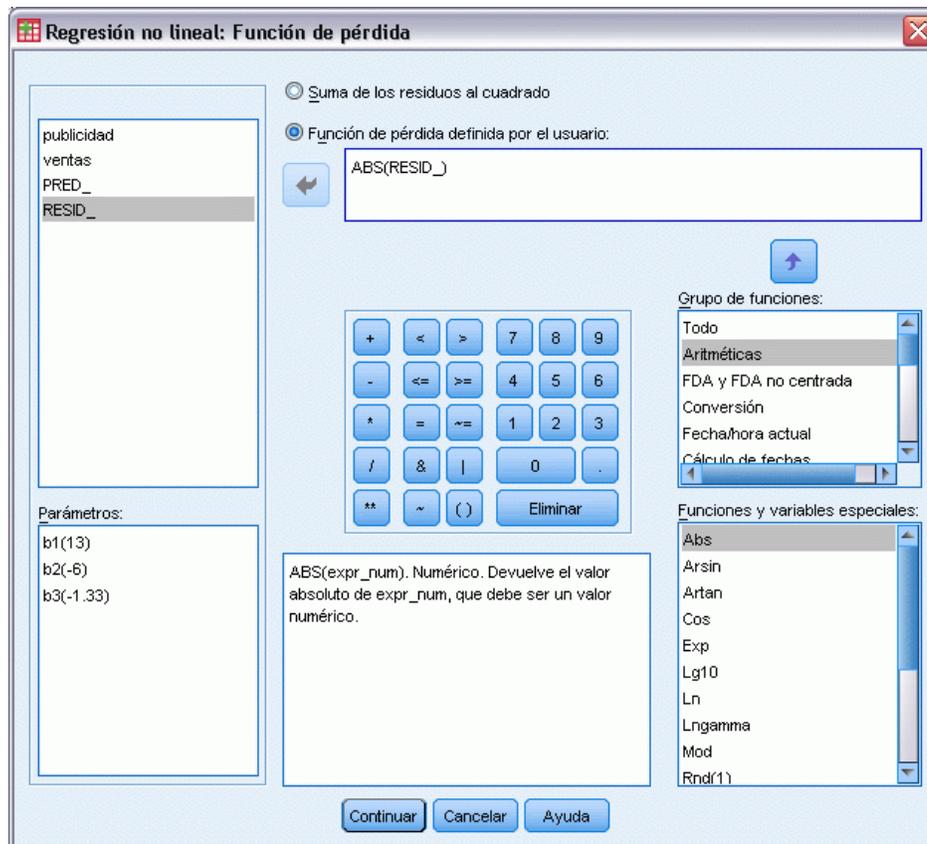
La tabla que se muestra más abajo proporciona ejemplos de la sintaxis del modelo para muchos modelos de regresión no lineal publicados. Es probable que los modelos seleccionados al azar no se ajusten bien a sus datos. Son necesarios valores iniciales adecuados para los parámetros; algunos modelos requieren restricciones para llegar a converger.

Tabla 5-1
Ejemplos de sintaxis de modelo

Nombre	Expresión del modelo
Regresión asintótica	$b1 + b2 * \exp(b3 * x)$
Regresión asintótica	$b1 - (b2 * (b3 ** x))$
Densidad	$(b1 + b2 * x)**(-1/b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
Log-modificado	$(b1 + b3 * x) ** b2$
Log-logístico	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Ley de Metcherlich de devoluciones decrecientes	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$
Razón de cúbicas	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
Razón de cuadráticas	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x ** b4)$
Densidad de rendimiento	$(b1 + b2 * x + b3 * x ** 2)**(-1)$

Regresión no lineal: Función de pérdida

Figura 5-3
Cuadro de diálogo Regresión no lineal: Función de pérdida



La **función de pérdida** en regresión no lineal es la función que es minimizada por el algoritmo. Seleccione Suma de los residuos al cuadrado para minimizar la suma de los residuos al cuadrado o bien Función de pérdida definida por el usuario para minimizar una función diferente.

Si selecciona Función de pérdida definida por el usuario, debe definir la función de pérdida cuya suma (en todos los casos) deberá minimizarse por la elección de los valores de los parámetros.

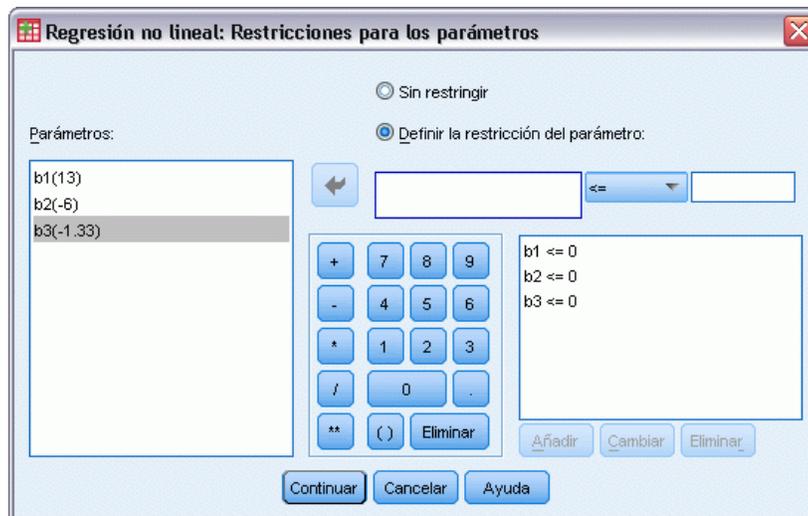
- La mayoría de las funciones de pérdida incluyen la variable especial *RESID_*, que representa el residuo. La función de pérdida por defecto Suma de los residuos al cuadrado se podría introducir explícitamente como $RESID_**2$. Si tiene que utilizar el valor pronosticado en la función de pérdida, éste es igual a la variable dependiente menos el residuo.
- Se puede especificar una función de pérdida condicional utilizando la lógica condicional.

Puede escribir una expresión en el campo Función de pérdida definida por el usuario, o bien pegar en el campo los componentes de la expresión. Las constantes de cadena deben ir entre comillas o apóstrofes y las constantes numéricas deben escribirse en formato americano, con el punto como separador de la parte decimal.

Regresión no lineal: Restricciones para los parámetros

Figura 5-4

Cuadro de diálogo Regresión no lineal: Restricciones para los parámetros



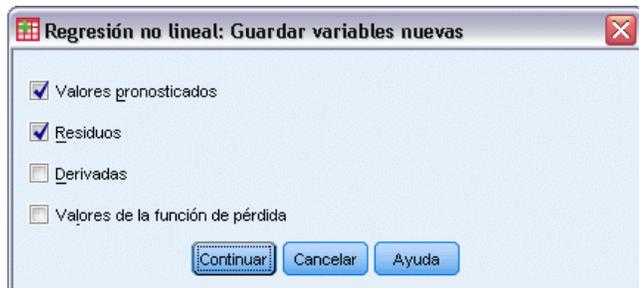
Una **restricción** es una limitación sobre los valores permitidos para un parámetro durante la búsqueda iterativa de una solución. Las expresiones lineales se evalúan antes de realizar un paso, de modo que se puedan utilizar restricciones lineales para omitir los pasos que pueden provocar desbordamientos. Las expresiones no lineales se evalúan después de realizar el paso.

Cada ecuación o desigualdad requiere los siguientes elementos:

- Una expresión que incluya al menos un parámetro del modelo. Escriba la expresión o bien utilice el teclado, que le permita pegar números, operadores o paréntesis en la expresión. Puede escribir el parámetro o parámetros requeridos junto con el resto de la expresión o bien pegarlos de la lista de Parámetros situada a la izquierda. No se pueden usar variables ordinarias en una restricción.
- Uno de los tres operadores lógicos \leq , $=$, o bien \geq .
- Una constante numérica, con la que se compara la expresión utilizando el operador lógico. Escriba la constante. Las constantes numéricas deben escribirse en formato americano, con el punto como separador de la parte decimal.

Regresión no lineal: Guardar variables nuevas

Figura 5-5
Cuadro de diálogo Regresión no lineal: Guardar variables nuevas

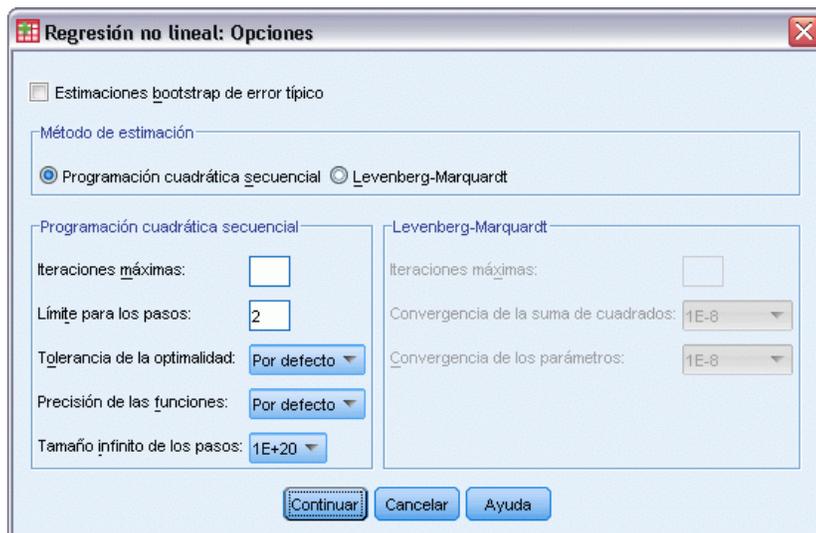


Puede guardar una serie de variables nuevas en el archivo de datos activo. Las opciones disponibles son: Residuos, Valores pronosticados, Derivadas y Valores de la función de pérdida. Estas variables se pueden utilizar en análisis subsiguientes para contrastar el ajuste del modelo o para identificar casos problemáticos.

- **Residuos.** Guarda los residuos con el nombre de variable resid.
- **Valores pronosticados.** Guarda los valores pronosticados, con el nombre de variable pred_.
- **Derivadas.** Se guarda una derivada para cada parámetro del modelo. Los nombres de las derivadas se construyen añadiendo como prefijo una "d." a los primeros seis caracteres de los nombres de los parámetros.
- **Valores de la función de pérdida.** Esta opción está disponible si especifica su propia función de pérdida. El nombre de variable loss_ está asignada a los valores de la función de pérdida.

Regresión no lineal: Opciones

Figura 5-6
Cuadro de diálogo Regresión no lineal: Opciones



Este cuadro de diálogo permite controlar diversos aspectos del análisis de regresión no lineal:

Estimaciones autodocimantes. Método para la estimación del error típico de un estadístico que usa muestras repetidas del conjunto original de datos. Se realiza mediante muestreo con repetición para obtener numerosas muestras del mismo tamaño que el conjunto de datos original. La ecuación no lineal se estima para cada una de estas muestras. A continuación, se calcula el error típico de cada estimación del parámetro como la desviación típica de las estimaciones creadas por el método autodocimante. Se utilizan los valores de los parámetros obtenidos a partir de los datos originales como los valores de inicio para las estimaciones autodocimantes. Requiere el algoritmo de programación cuadrática secuencial.

Método de estimación. Permite seleccionar un método de estimación, si esto es posible. (Determinadas opciones de éste y otros cuadros de diálogo requieren el algoritmo de programación cuadrática secuencial.) Entre las alternativas disponibles se encuentran Programación cuadrática secuencial y Levenberg-Marquardt.

- **Programación cuadrática secuencial.** Este método está disponible para los modelos restringidos y los no restringidos. La programación cuadrática secuencial se utiliza automáticamente si especifica un modelo restringido, una función de pérdida definida por el usuario o el muestreo bootstrap. Puede introducir nuevos valores para N° máximo de iteraciones y Límite para los pasos, puede cambiar la selección en las listas desplegables de Tolerancia de la optimidad, Precisión de la función y Tamaño para pasos infinitos.
- **Levenberg-Marquardt.** Algoritmo por defecto para los modelos no restringidos. No puede utilizar el método de Levenberg-Marquardt si especifica un modelo restringido, una función de pérdida definida por el usuario, o el muestreo autodocimante. Se pueden introducir nuevos valores para N° máximo de iteraciones y cambiar las selecciones de las listas desplegables de Convergencia de la suma de cuadrados y Convergencia en los parámetros.

Regresión no lineal: Interpretación de los resultados

Los problemas de la regresión no lineal presentan con frecuencia dificultades de cálculo:

- La elección de valores iniciales para los parámetros influye en la convergencia. Intente elegir valores iniciales que sean razonables y, si es posible, próximos a la solución final esperada.
- A veces, un algoritmo se ejecuta mejor que otro en un determinado problema. En el cuadro de diálogo Opciones, seleccione el otro algoritmo si está disponible. Si especifica una función de pérdida o ciertos tipos de restricciones, no podrá utilizar el algoritmo de Levenberg-Marquardt.
- Cuando la iteración se detiene sólo porque se ha dado el número máximo de iteraciones, el modelo «final» probablemente no sea una buena solución. Seleccione Usar los valores iniciales del análisis previo del cuadro de diálogo Parámetros para continuar la iteración o, mejor aún, elija otros valores iniciales.
- Los modelos que requieren la exponenciación de o por valores de datos grandes pueden causar desbordamientos (números demasiado grandes o demasiado pequeños para que su equipo los represente). A veces esto se puede evitar eligiendo los valores iniciales adecuados o imponiendo restricciones sobre los parámetros.

Funciones adicionales del comando NLR

Con el lenguaje de sintaxis de comandos también podrá:

- Nombrar un archivo del cual leer valores iniciales para las estimaciones de los parámetros.
- Especificar más de una instrucción de modelo y de función de pérdida. Con ello se facilita la tarea de especificación de un modelo segmentado.
- Suministrar sus propias derivadas en vez de utilizar las calculadas por el programa.
- Especificar el número de muestras bootstrap que se van a generar.
- Especificar criterios de iteración adicionales, incluyendo el establecer un valor crítico para la comprobación de las derivadas y definir un criterio de convergencia para la correlación entre los residuos y las derivadas.

Los criterios adicionales para el comando CNLR (regresión no lineal restringida) permiten:

- Especificar el número máximo de iteraciones secundarias permitidas dentro de cada iteración principal.
- Establecer un valor crítico para la comprobación de las derivadas.
- Establecer un límite para los pasos.
- Especificar una tolerancia de colapso para determinar si los valores iniciales están dentro de los límites especificados.

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Estimación ponderada

Los modelos de regresión lineal típicos asumen que la varianza es constante en la población objeto de estudio. Cuando éste no es el caso (por ejemplo cuando los casos con puntuaciones mayores en un atributo muestran más variabilidad que los casos con puntuaciones menores en ese atributo), la regresión lineal mediante mínimos cuadrados ordinarios (MCO, OLS) deja de proporcionar estimaciones óptimas para el modelo. Si las diferencias de variabilidad se pueden pronosticar a partir de otra variable, el procedimiento Estimación ponderada permite calcular los coeficientes de un modelo de regresión lineal mediante mínimos cuadrados ponderados (MCP, WLS), de forma que se les dé mayor ponderación a las observaciones más precisas (es decir, aquellas con menos variabilidad) al determinar los coeficientes de regresión. El procedimiento Estimación ponderada contrasta un rango de transformaciones de ponderación e indica cuál se ajustará mejor a los datos.

Ejemplo. ¿Cuáles son los efectos de la inflación y el paro sobre los cambios en el precio de las acciones Debido a que los valores con mayor valor de cotización suelen mostrar más variabilidad que aquellos con menor valor de cotización, la estimación de mínimos cuadrados ordinarios no generará estimaciones que sean óptimas. El método de Estimación ponderada permite capturar el efecto del precio de cotización sobre la variabilidad de los cambios en el precio, al calcular el modelo lineal.

Estadísticos. Valores de la log-verosimilitud para cada potencia de la variable de ponderación puesta a prueba, R múltiple, R -cuadrado, R cuadrado corregida, tabla de ANOVA para el modelo MCP, estimaciones de los parámetros tipificados y no tipificados y log-verosimilitud para el modelo MCP.

Datos. Las variables dependiente e independientes deben ser cuantitativas. Las variables categóricas, como la religión, la mayoría de edad o el lugar de residencia, han de recodificarse como variables binarias (dummy) o como otro de los tipos de variables de contraste. La variable de ponderación deberá ser cuantitativa y estar relacionada con la variabilidad de la variable dependiente.

Supuestos. Para cada valor de la variable independiente, la distribución de la variable dependiente debe ser normal. La relación entre la variable dependiente y cada variable independiente debe ser lineal y todas las observaciones deben ser independientes. La varianza de la variable dependiente puede cambiar según los niveles de la variable o variables independientes, pero las diferencias se deben poder pronosticar en función de la variable de ponderación.

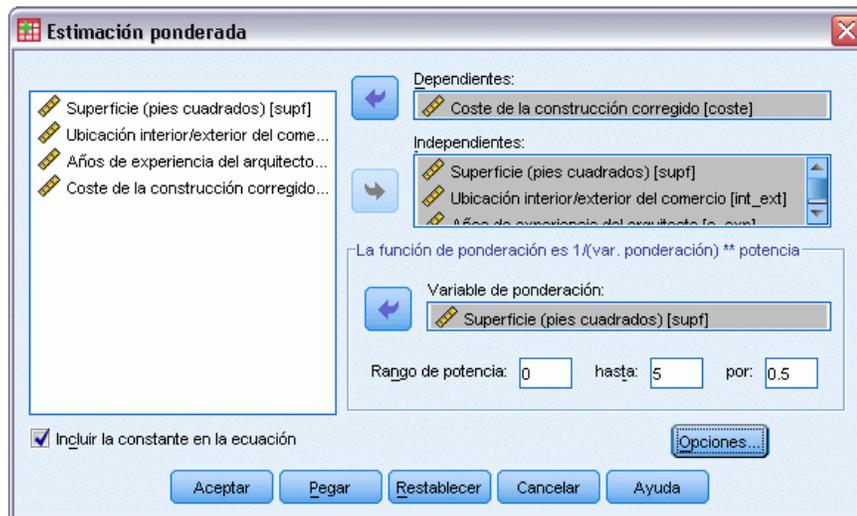
Procedimientos relacionados. El procedimiento Explorar se puede utilizar para inspeccionar los datos. Este procedimiento proporciona pruebas de normalidad y homogeneidad de la varianza, así como representaciones gráficas. Si la variable dependiente parece tener la misma varianza para todos los niveles de las variables independientes, puede utilizar el procedimiento Regresión lineal. Si los datos parecen violar un supuesto (como puede ser la normalidad), intente transformarlos. Si los datos no están relacionados linealmente y una transformación no ayuda, utilice un

modelo alternativo en el procedimiento Estimación curvilínea. Si la variable dependiente es dicotómica, por ejemplo si se lleva o no a cabo una determinada venta o si un artículo es o no defectuoso, utilice el procedimiento Regresión logística. Si la variable dependiente está censurada (por ejemplo, el tiempo de supervivencia después de una intervención quirúrgica), utilice los procedimientos Tablas de mortalidad, Kaplan-Meier o Regresión de Cox, disponibles en la opción Estadísticas avanzadas. Si los datos no son independientes (por ejemplo, si observa a la misma persona en diversas condiciones), utilice el procedimiento Medidas repetidas, disponible en la opción Estadísticas avanzadas.

Para obtener un análisis de Estimación ponderada

- ▶ Elija en los menús:
Analizar > Regresión > Estimación ponderada...

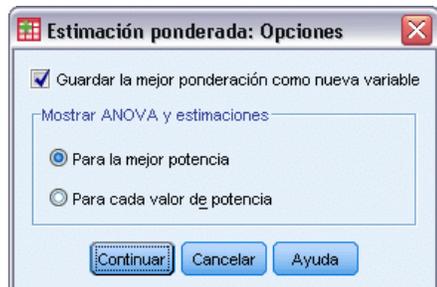
Figura 6-1
Cuadro de diálogo Estimación ponderada



- ▶ Seleccione una variable dependiente.
- ▶ Seleccionar una o más variables independientes.
- ▶ Seleccione la variable fuente de heterocedasticidad como variable de ponderación.
 - **Variable de ponderación.** Los datos son ponderados por el inverso de esta variable elevada a una potencia. La ecuación de regresión se calcula para cada uno de los valores del rango de potencias especificado y se indica la potencia que maximiza el logaritmo de la función de verosimilitud.
 - **Rango de potencias.** Se utiliza en conjunción con la variable de ponderación para calcular las ponderaciones. Se ajustarán varias ecuaciones de regresión, una para cada valor del rango de potencias. Los valores introducidos en el cuadro Rango de potencias y en el cuadro Hasta, deben encontrarse entre -6,5 y 7,5, ambos inclusive. Los valores de la potencia irán del valor menor al mayor, con el incremento determinado por el valor especificado. El número total de valores del rango de potencias está limitado a 150.

Estimación ponderada: Opciones

Figura 6-2
Cuadro de diálogo Estimación ponderada: Opciones



Puede especificar las opciones para el análisis de estimación ponderada:

Guardar la mejor ponderación como nueva variable. Añade la variable de ponderación al archivo activo. Esta variable se denomina *WGT_n*, siendo *n* un número elegido para proporcionar a la variable un nombre exclusivo.

Mostrar ANOVA y estimaciones. Permite controlar cómo se muestran los estadísticos en los resultados. Las alternativas disponibles son Para la mejor potencia y Para cada valor de potencia.

Funciones adicionales del comando WLS

Con el lenguaje de sintaxis de comandos también podrá:

- Proporcionar un único valor para la potencia.
- Especificar una lista de valores para la potencia o mezclar un rango de valores con una lista de valores para la potencia.

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Regresión por mínimos cuadrados en dos fases

Los modelos de regresión lineal típica asumen que los errores de la variable dependiente no están correlacionados con la variable o variables independientes. Cuando éste no es el caso (por ejemplo, cuando las relaciones entre las variables son bidireccionales), la regresión lineal mediante mínimos cuadrados ordinarios (OLS) deja de proporcionar estimaciones óptimas del modelo. La regresión por mínimos cuadrados en dos fases utiliza variables instrumentales que no estén correlacionadas con los términos de error para calcular los valores estimados de los predictores problemáticos (en la primera fase) y después utiliza dichos valores calculados para estimar un modelo de regresión lineal para la variable dependiente (la segunda fase). Dado que los valores calculados se basan en variables que no están correlacionadas con los errores, los resultados del modelo en dos fases son óptimos.

Ejemplo. ¿Está relacionada la demanda de un artículo con su precio y con los ingresos del consumidor? La dificultad de este modelo radica en que el precio y la demanda tienen efectos recíprocos el uno sobre el otro. Es decir, el precio puede influir en la demanda y la demanda también puede influir en el precio. Un modelo de regresión por mínimos cuadrados en dos fases permite utilizar los ingresos de los consumidores y el precio retardado para calcular un predictor sustituto del precio, el cual no esté correlacionado con los errores de medida de la demanda. Se reemplaza el precio en el modelo especificado originariamente por este sustituto y después se estima el nuevo modelo.

Estadísticos. Para cada modelo: coeficientes de regresión tipificados y no tipificados, R múltiple, R^2 cuadrado, R^2 cuadrado corregida, error típico de la estimación, tabla de análisis de varianza, valores pronosticados y residuos. Además, los intervalos de confianza al 95% para cada coeficiente de regresión y las matrices de correlación y covarianza para las estimaciones de los parámetros.

Datos. Las variables dependiente e independientes deben ser cuantitativas. Las variables categóricas, como la religión, la mayoría de edad o el lugar de residencia, han de recodificarse como variables binarias (dummy) o como otro de los tipos de variables de contraste. Las variables explicativas **endógenas** deben ser cuantitativas (no categóricas).

Supuestos. Para cada valor de la variable independiente, la distribución de la variable dependiente debe ser normal. La varianza de distribución de la variable dependiente debe ser constante para todos los valores de la variable independiente. La relación entre la variable dependiente y cada variable independiente debe ser lineal.

Procedimientos relacionados. Si piensa que ninguna de las variables predictoras está correlacionada con los errores de la variable dependiente, puede utilizar el procedimiento Regresión lineal. Si los datos parecen violar alguno de los supuestos (como la normalidad o la varianza constante), pruebe a transformarlos. Si los datos no están relacionados linealmente y una transformación no

ayuda, utilice un modelo alternativo en el procedimiento Estimación curvilínea. Si la variable dependiente es dicotómica, por ejemplo si se ha completado o no una determinada venta, utilice el procedimiento Regresión logística. Si los datos no son independientes (por ejemplo, si observa a la misma persona en diversas condiciones), utilice el procedimiento Medidas repetidas, disponible en la opción Modelos avanzados.

Para obtener un análisis de regresión de mínimos cuadrados en dos fases

- ▶ En los menús, seleccione:
Analizar > Regresión > Mínimos cuadrados en dos fases...

Figura 7-1
Cuadro de diálogo Mínimos cuadrados en dos fases



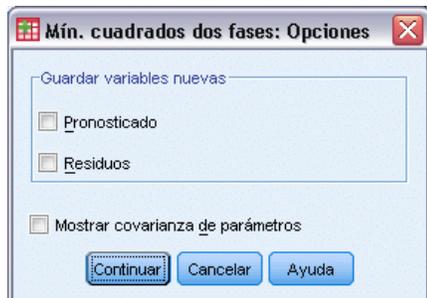
- ▶ Seleccione una variable dependiente.
- ▶ Seleccione una o más variables (predictoras) explicativas.
- ▶ Seleccione una o más variables instrumentales.
 - **Instrumental.** Estas son las variables utilizadas para calcular los valores pronosticados de las variables endógenas en la primera etapa del análisis de mínimos cuadrados en dos fases. Las listas de variables explicativas e instrumentales pueden contener las mismas variables. Debe haber al menos tantas variables instrumentales como explicativas. Si las variables explicativas e instrumentales coinciden, los resultados serán los mismos que los del procedimiento de Regresión lineal.

Las variables explicativas no especificadas como instrumentales se consideran endógenas. Normalmente, todas las variables exógenas de la lista Explicativas se especifican también como variables instrumentales.

Regresión por mínimos cuadrados en dos fases: Opciones

Figura 7-2

Cuadro de diálogo Mínimos cuadrados en dos fases: Opciones



Puede seleccionar las opciones siguientes para su análisis:

Guardar variables nuevas. Permite añadir nuevas variables al archivo activo. Las opciones disponibles son Pronósticos y Residuos.

Mostrar covarianza de parámetros. Permite imprimir la matriz de covarianzas de las estimaciones de los parámetros.

Funciones adicionales del comando 2SLS

El lenguaje de sintaxis de comandos también permite estimar varias ecuaciones simultáneamente. Consulte *Command Syntax Reference* para una referencia completa de sintaxis de comandos.

Esquemas de codificación de variables categóricas

En muchos procedimientos, se puede solicitar la sustitución automática de una variable independiente categórica por un conjunto de variables de contraste, que se podrán introducir o eliminar de una ecuación como un bloque. Puede especificar cómo se va a codificar el conjunto de variables de contraste, normalmente en el subcomando `CONTRAST`. Ese apéndice explica e ilustra el funcionamiento real de los distintos tipos de contrastes solicitados en `CONTRAST`.

Desviación

Desviación desde la media global. En términos matriciales, estos contrastes tienen la forma:

media	($1/k$	$1/k$...	$1/k$	$1/k$)
gl(1)	($1-1/k$	$-1/k$...	$-1/k$	$-1/k$)
gl(2)	($-1/k$	$1-1/k$...	$-1/k$	$-1/k$)
.			.				
.			.				
gl(k-1)	($-1/k$	$-1/k$...	$1-1/k$	$-1/k$)

donde k es el número de categorías para la variable independiente y, por defecto, se omite la última categoría. Por ejemplo, los contrastes de desviación para una variable independiente con tres categorías son los siguientes:

($1/3$	$1/3$	$1/3$)
($2/3$	$-1/3$	$-1/3$)
($-1/3$	$2/3$	$-1/3$)

Para omitir una categoría distinta de la última, especifique el número de la categoría omitida entre el paréntesis que sucede a la palabra clave `DEVIATION`. Por ejemplo, el siguiente subcomando obtiene las desviaciones para la primera y tercera categorías y omite la segunda:

```
/CONTRAST (FACTOR) =DEVIATION (2)
```

Suponga que *factor* tiene tres categorías. La matriz de contraste resultante será

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & -1/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$$

Simple

Contrastes simples. Compara cada nivel de un factor con el último. La forma de la matriz general es

$$\begin{array}{l} \text{media} \\ \text{gl(1)} \\ \text{gl(2)} \\ \vdots \\ \text{gl(k-1)} \end{array} \begin{pmatrix} 1/k & 1/k & \dots & 1/k & 1/k \\ 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & & & \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

donde k es el número de categorías para la variable independiente. Por ejemplo, los contrastes simples para una variable independiente con cuatro categorías son los siguientes:

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Para utilizar otra categoría en lugar de la última como categoría de referencia, especifique entre paréntesis tras la palabra clave `SIMPLE` el número de secuencia de la categoría de referencia, que no es necesariamente el valor asociado con dicha categoría. Por ejemplo, el siguiente subcomando `CONTRAST` obtiene una matriz de contraste que omite la segunda categoría:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Suponga que *factor* tiene cuatro categorías. La matriz de contraste resultante será

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

Helmert

Contrastes de Helmert. Compara categorías de una variable independiente con la media de las categorías subsiguientes. La forma de la matriz general es

$$\begin{array}{l}
 \text{media} \quad (1/k \quad 1/k \quad \dots \quad 1/k \quad 1/k) \\
 \text{gl(1)} \quad (1 \quad -1/(k-1) \quad \dots \quad -1/(k-1) \quad -1/(k-1)) \\
 \text{gl(2)} \quad (0 \quad 1 \quad \dots \quad -1/(k-2) \quad -1/(k-2)) \\
 \cdot \\
 \cdot \\
 \text{gl}(k-2) \quad (0 \quad 0 \quad 1 \quad -1/2 \quad -1/2) \\
 \text{gl}(k-1) \quad (0 \quad 0 \quad \dots \quad 1 \quad -1)
 \end{array}$$

donde k es el número de categorías de la variable independiente. Por ejemplo, una variable independiente con cuatro categorías tiene una matriz de contraste de Helmert con la siguiente forma:

$$\begin{array}{l}
 (1/4 \quad 1/4 \quad 1/4 \quad 1/4) \\
 (1 \quad -1/3 \quad -1/3 \quad -1/3) \\
 (0 \quad 1 \quad -1/2 \quad -1/2) \\
 (0 \quad 0 \quad 1 \quad -1)
 \end{array}$$

Diferencia

Diferencia o contrastes de Helmert inversos. Compara categorías de una variable independiente con la media de las categorías anteriores de la variable. La forma de la matriz general es

$$\begin{array}{l}
 \text{media} \quad (1/k \quad 1/k \quad 1/k \quad \dots \quad 1/k) \\
 \text{gl(1)} \quad (-1 \quad 1 \quad 0 \quad \dots \quad 0) \\
 \text{gl(2)} \quad (-1/2 \quad -1/2 \quad 1 \quad \dots \quad 0) \\
 \cdot \\
 \cdot \\
 \text{gl}(k-1) \quad (-1/(k-1) \quad -1/(k-1) \quad -1/(k-1) \quad \dots \quad 1)
 \end{array}$$

donde k es el número de categorías para la variable independiente. Por ejemplo, los contrastes de diferencia para una variable independiente con cuatro categorías son los siguientes:

$$\begin{array}{l}
 (1/4 \quad 1/4 \quad 1/4 \quad 1/4) \\
 (-1 \quad 1 \quad 0 \quad 0) \\
 (-1/2 \quad -1/2 \quad 1 \quad 0) \\
 (-1/3 \quad -1/3 \quad -1/3 \quad 1)
 \end{array}$$

Polinómico

Contrastes polinómicos ortogonales. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, el tercer grado de libertad, el cúbico, y así sucesivamente hasta los efectos de orden superior.

Se puede especificar el espaciado entre niveles del tratamiento medido por la variable categórica dada. Se puede especificar un espaciado igual, que es el valor por defecto si se omite la métrica, como enteros consecutivos desde 1 hasta k , donde k es el número de categorías. Si la variable *fármaco* tiene tres categorías, el subcomando

```
/CONTRAST (DRUG) = POLYNOMIAL
```

es idéntico a

```
/CONTRAST (DRUG) = POLYNOMIAL (1, 2, 3)
```

De todas maneras, el espaciado igual no es siempre necesario. Por ejemplo, supongamos que *fármaco* representa las diferentes dosis de un fármaco administrado a tres grupos. Si la dosis administrada al segundo grupo es el doble que la administrada al primer grupo y la dosis administrada al tercer grupo es el triple que la del primer grupo, las categorías del tratamiento están espaciadas por igual y una métrica adecuada para esta situación se compone de enteros consecutivos:

```
/CONTRAST (DRUG) = POLYNOMIAL (1, 2, 3)
```

No obstante, si la dosis administrada al segundo grupo es cuatro veces la administrada al primer grupo, y la dosis del tercer grupo es siete veces la del primer grupo, una métrica adecuada sería

```
/CONTRAST (DRUG) = POLYNOMIAL (1, 4, 7)
```

En cualquier caso, el resultado de la especificación de contrastes es que el primer grado de libertad para *fármaco* contiene el efecto lineal de los niveles de dosificación y el segundo grado de libertad contiene el efecto cuadrático.

Los contrastes polinómicos son especialmente útiles en contrastes de tendencias y para investigar la naturaleza de superficies de respuestas. También se pueden utilizar contrastes polinómicos para realizar ajustes de curvas no lineales, como una regresión curvilínea.

Repetido

Compara niveles adyacentes de una variable independiente. La forma de la matriz general es

media	(1/k	1/k	1/k	...	1/k	1/k)
gl(1)	(1	-1	0	...	0	0)
gl(2)	(0	1	-1	...	0	0)
.
.
gl(k-1)	(0	0	0	...	1	-1)

donde k es el número de categorías para la variable independiente. Por ejemplo, los contrastes repetidos para una variable independiente con cuatro categorías son los siguientes:

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Estos contrastes son útiles en el análisis de perfiles y siempre que sean necesarias puntuaciones de diferencia.

Especial

Un contraste definido por el usuario. Permite la introducción de contrastes especiales en forma de matrices cuadradas con tantas filas y columnas como categorías haya de la variable independiente. Para MANOVA y LOGLINEAR, la primera fila introducida es siempre el efecto promedio, o constante, y representa el conjunto de ponderaciones que indican cómo promediar las demás variables independientes, si las hay, sobre la variable dada. Generalmente, este contraste es un vector de contrastes.

Las restantes filas de la matriz contienen los contrastes especiales que indican las comparaciones deseadas entre categorías de la variable. Normalmente, los contrastes ortogonales son los más útiles. Este tipo de contrastes son estadísticamente independientes y son no redundantes. Los contrastes son ortogonales si:

- Para cada fila, la suma de los coeficientes de contrastes es igual a cero.
- Los productos de los correspondientes coeficientes para todos los pares de filas disjuntas también suman cero.

Por ejemplo, supongamos que el tratamiento tiene cuatro niveles y que deseamos comparar los diversos niveles del tratamiento entre sí. Un contraste especial adecuado sería

$(1 \quad 1 \quad 1 \quad 1)$	ponderaciones para el cálculo de la media
$(3 \quad -1 \quad -1 \quad -1)$	compare 1° con 2° hasta 4°
$(0 \quad 2 \quad -1 \quad -1)$	compare 2° con 3° y 4°
$(0 \quad 0 \quad 1 \quad -1)$	compare 3° con 4°

todo lo cual se especifica mediante el siguiente subcomando CONTRAST para MANOVA, LOGISTIC REGRESSION y COXREG:

```
/CONTRAST (TREATMNT) =SPECIAL ( 1 1 1 1
                                3 -1 -1 -1
                                0 2 -1 -1
                                0 0 1 -1 )
```

Para LOGLINEAR, es necesario especificar:

```
/CONTRAST (TREATMNT) =BASIS SPECIAL ( 1 1 1 1
                                       3 -1 -1 -1
                                       0 2 -1 -1
```

$$0 \ 0 \ 1 \ -1 \)$$

Cada fila, excepto la fila de las medias suman cero. Los productos de cada par de filas disjuntas también suman cero:

$$\text{Filas 2 y 3:} \quad (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0$$

$$\text{Filas 2 y 4:} \quad (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0$$

$$\text{Filas 3 y 4:} \quad (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0$$

No es necesario que los contrastes especiales sean ortogonales. No obstante, no deben ser combinaciones lineales de unos con otros. Si lo son, el procedimiento informará de la dependencia lineal y detendrá el procesamiento. Los contrastes de Helmert, de diferencia y polinómicos son todos contrastes ortogonales.

Indicador

Codificación de la variable indicadora. También conocida como variable auxiliar o dummy, no está disponible en LOGLINEAR o MANOVA. El número de variables nuevas codificadas es $k-1$. Los casos que pertenezcan a la categoría de referencia se codificarán como 0 para las k -variables 1. Un caso en la categoría i ésima se codificará como 0 para todas las variables indicadoras excepto la i ésima, que se codificará como 1.

Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing,

using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



- Análisis probit
 - critérios, 24
 - definición de rango, 24
 - ejemplo, 22
 - estadísticos, 22, 24
 - funciones adicionales del comando, 25
 - intervalos de confianza fiduciaria, 24
 - iteraciones, 24
 - potencia relativa de la mediana, 24
 - prueba de paralelismo, 24
 - tasa de respuesta natural, 24
- bondad de ajuste
 - en Regresión logística multinomial, 16
- casillas con 0 observaciones
 - en Regresión logística multinomial, 18
- categoría de referencia
 - en Regresión logística multinomial, 15
- chi-cuadrado de Pearson
 - bondad de ajuste, 16
 - estimación del valor del escalamiento para la dispersión, 19
- clasificación
 - en Regresión logística multinomial, 11
- contrastes
 - en Regresión logística, 6
- covariables
 - en Regresión logística, 6
- covariables categóricas, 6
- covariables de cadena
 - en Regresión logística, 6
- criterio de convergencia
 - en Regresión logística multinomial, 18
- D de Cook
 - en Regresión logística, 7
- delta
 - corrección para casillas con 0 observaciones, 18
- DfBeta
 - en Regresión logística, 7
- eliminación hacia atrás
 - en Regresión logística, 5
- estadístico de bondad de ajuste de Hosmer-Lemeshow
 - en Regresión logística, 9
- Estimación ponderada, 35
 - almacenamiento de la mejor ponderación como una variable nueva, 37
 - ejemplo, 35
 - estadísticos, 35
 - funciones adicionales del comando, 37
 - historial de iteraciones, 37
 - log de la verosimilitud, 35
 - mostrar ANOVA y estimaciones, 37
- estimaciones de los parámetros
 - en Regresión logística multinomial, 16
- función de desviación
 - estimación del valor del escalamiento para la dispersión, 19
- histórico de iteraciones
 - en Regresión logística multinomial, 18
- intersección
 - incluir o excluir, 13
- intervalos de confianza
 - en Regresión logística multinomial, 16
- intervalos de confianza fiduciaria
 - en el análisis probit, 24
- iteraciones
 - en el análisis probit, 24
 - en Regresión logística, 9
 - en Regresión logística multinomial, 18
- legal notices, 47
- ley de Metcherlich de devoluciones decrecientes
 - en Regresión no lineal, 29
- log de la verosimilitud
 - en Estimación ponderada, 35
 - en Regresión logística multinomial, 16
- matriz de correlaciones
 - en Regresión logística multinomial, 16
- matriz de covarianzas
 - en Regresión logística multinomial, 16
- modelo de densidad
 - en Regresión no lineal, 29
- modelo de densidad de rendimiento
 - en Regresión no lineal, 29
- modelo de Gauss
 - en Regresión no lineal, 29
- modelo de Gompertz
 - en Regresión no lineal, 29
- modelo de Johnson-Schumacher
 - en Regresión no lineal, 29
- modelo de Michaelis Menten
 - en Regresión no lineal, 29

- modelo de Morgan-Mercer-Florin
 - en Regresión no lineal, 29
- modelo de Peal-Reed
 - en Regresión no lineal, 29
- modelo de razón de cuadráticas
 - en Regresión no lineal, 29
- modelo de razón de cúbicas
 - en Regresión no lineal, 29
- modelo de Richards
 - en Regresión no lineal, 29
- modelo de Verhulst
 - en Regresión no lineal, 29
- modelo de Von Bertalanffy
 - en Regresión no lineal, 29
- modelo de Weibull
 - en Regresión no lineal, 29
- modelo log-modificado
 - en Regresión no lineal, 29
- modelos de efectos principales
 - en Regresión logística multinomial, 13
- modelos factoriales completos
 - en Regresión logística multinomial, 13
- modelos no lineales
 - en Regresión no lineal, 29
- modelos personalizados
 - en Regresión logística multinomial, 13

- potencia relativa de la mediana
 - en el análisis probit, 24
- prueba de paralelismo
 - en el análisis probit, 24

- R cuadrado de McFadden
 - en Regresión logística multinomial, 16
- R-cuadrado de Cox y Snell
 - en Regresión logística multinomial, 16
- R-cuadrado de Nagelkerke
 - en Regresión logística multinomial, 16
- razón de verosimilitud
 - bondad de ajuste, 16
 - estimación del valor del escalamiento para la dispersión, 19
- regresión asintótica
 - en Regresión no lineal, 29
- Regresión lineal
 - estimación ponderada, 35
 - Regresión por mínimos cuadrados en dos fases, 38
- Regresión logística
 - binaria, 1
- Regresión Logística, 3
 - almacenamiento de nuevas variables, 7
 - coeficientes, 3
 - contrastos, 6
 - covariables categóricas, 6
 - covariables de cadena, 6
 - definición de la regla de selección, 5
 - ejemplo, 3
 - establecimiento de reglas, 5
 - estadístico de bondad de ajuste de Hosmer-Lemeshow, 9
 - estadísticos, 3
 - estadísticos y gráficos, 9
 - funciones adicionales del comando, 10
 - iteraciones, 9
 - medidas de influencia, 7
 - métodos de selección de variables, 5
 - opciones de representación, 9
 - probabilidad para el método por pasos, 9
 - punto de corte para la clasificación, 9
 - residuos, 7
 - término constante, 9
 - valores pronosticados, 7
- regresión logística binaria, 1
- Regresión logística multinomial, 11, 16
 - categoría de referencia, 15
 - criterios, 18
 - estadísticos, 16
 - exportación de información del modelo, 20
 - funciones adicionales del comando, 21
 - guardar, 20
 - modelos, 13
- Regresión no lineal, 26
 - algoritmo de Levenberg-Marquardt, 32
 - almacenamiento de variables nuevas, 32
 - derivadas, 32
 - ejemplo, 26
 - estadísticos, 26
 - estimaciones bootstrap, 32
 - función de pérdida, 30
 - funciones adicionales del comando, 34
 - interpretación de resultados, 33
 - lógica condicional, 27
 - métodos de estimación, 32
 - modelo segmentado, 27
 - modelos no lineales comunes, 29
 - parámetros, 28
 - programación cuadrática secuencial, 32
 - residuos, 32
 - restricciones para los parámetros, 31
 - valores iniciales, 28
 - valores pronosticados, 32
- Regresión por mínimos cuadrados en dos fases, 38
 - almacenamiento de nuevas variables, 40
 - covarianzas de los parámetros, 40
 - ejemplo, 38
 - estadísticos, 38
 - funciones adicionales del comando, 40
 - variables instrumentales, 38
- regresión restringida
 - en Regresión no lineal, 31
- restricciones para los parámetros
 - en Regresión no lineal, 31

- selección hacia delante
 - en Regresión logística, 5

-
- selección por pasos
 - en Regresión logística, 5
 - en Regresión logística multinomial, 13
 - separación
 - en Regresión logística multinomial, 18
 - singularidad
 - en Regresión logística multinomial, 18
 - subdivisión por pasos
 - en Regresión logística multinomial, 18

 - tablas de clasificación
 - en Regresión logística multinomial, 16
 - tablas de probabilidades de casilla
 - en Regresión logística multinomial, 16
 - término constante
 - en Regresión lineal, 9
 - trademarks, 48

 - valor del escalamiento para la dispersión
 - en Regresión logística multinomial, 19
 - valores de influencia
 - en Regresión logística, 7