# IBM SPSS Advanced Statistics 19

*Note*: Before using this information and the product it supports, read the general information under Notices el p. 170.

# Prefacio

IBM® SPSS® Statistics es un sistema global para el análisis de datos. El módulo adicional opcional Estadísticas avanzadas proporciona las técnicas de análisis adicionales que se describen en este manual. El módulo adicional Estadísticas avanzadas 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 Introducción a Estadísticas avanzadas 1

## 2 Análisis MLG multivariante 2

Modelo MLG multivariante . . . . . 4
Construir términos . . . . . 5
Suma de cuadrados . . . . . 5
MLG Multivariante: Contrastes . . . . . 7
Tipos de contrastes . . . . . 7
Gráficos de perfil de MLG multivariante . . . . . 8
MLG multivariante: Comparaciones post hoc . . . . . 9
MLG: Guardar . . . . . 11
MLG Multivariante: Opciones . . . . . 12
Funciones adicionales del comando GLM . . . . . 14

## 3 MLG Medidas repetidas 15

MLG Medidas repetidas: Definir factores . . . . . 18
MLG Medidas repetidas: Modelo . . . . . 20
Construir términos . . . . . 21
Suma de cuadrados . . . . . 21
MLG Medidas repetidas: Contrastes . . . . . 22
Tipos de contrastes . . . . . 23
MLG Medidas repetidas: Gráficos de perfil . . . . . 23
MLG Medidas repetidas: Comparaciones post hoc . . . . . 24
MLG medidas repetidas: Guardar . . . . . 26
MLG Medidas repetidas: Opciones . . . . . 28
Funciones adicionales del comando GLM . . . . . 29

## 4 Análisis de componentes de la varianza 31

Componentes de la varianza: Modelo . . . . . 33
Construir términos . . . . . 33

Componentes de la varianza: Opciones . . . . . 34
Sumas de cuadrados (Componentes de la varianza). . . . . 35
Componentes de la varianza: Guardar en archivo nuevo . . . . . 36
Funciones adicionales del comando VARCOMP . . . . . 36

## 5 Modelos lineales mixtos 37

Modelos lineales mixtos: Selección de las variables de Sujetos/Repetidas . . . . . 39
Efectos fijos de los Modelos lineales mixtos . . . . . 41
Construir términos no anidados . . . . . 41
Construir términos anidados . . . . . 42
Suma de cuadrados. . . . . 42
Efectos aleatorios de los Modelos lineales mixtos . . . . . 43
Estimación de los Modelos lineales mixtos . . . . . 45
Estadísticos de Modelos lineales mixtos . . . . . 46
Medias marginales estimadas de modelos lineales mixtos . . . . . 47
Guardar Modelos lineales mixtos . . . . . 48
Funciones adicionales del comando MIXED . . . . . 49

## 6 Modelos lineales generalizados 50

Modelos lineales generalizados: Respuesta . . . . . 55
Modelos lineales generalizados: Categoría de referencia . . . . . 56
Modelos lineales generalizados: Predictores . . . . . 57
Modelos lineales generalizados: Opciones . . . . . 58
Modelos lineales generalizados: Modelo . . . . . 59
Modelos lineales generalizados: Estimación. . . . . 61
Modelos lineales generalizados: Valores iniciales . . . . . 63
Modelos lineales generalizados: Estadísticos . . . . . 64
Modelos lineales generalizados: Medias marginales estimadas . . . . . 66
Modelos lineales generalizados: Guardar . . . . . 68
Modelos lineales generalizados: Exportar. . . . . 70
Funciones adicionales del comando GENLIN . . . . . 71

## 7 Ecuaciones de estimación generalizadas 73

Ecuaciones de estimación generalizadas: Tipo de modelo . . . . . 77
Respuesta de las ecuaciones de estimación generalizadas . . . . . 80
Ecuaciones de estimación generalizadas: Categoría de referencia . . . . . 81
Ecuaciones de estimación generalizadas: Predictores . . . . . 83
Ecuaciones de estimación generalizadas: Opciones . . . . . 84
Ecuaciones de estimación generalizadas: Modelo . . . . . 85
Ecuaciones de estimación generalizadas: Estimación. . . . . 87
Ecuaciones de estimación generalizadas: Valores iniciales . . . . . 89
Ecuaciones de estimación generalizadas: Estadísticos . . . . . 91
Ecuaciones de estimación generalizadas: Medias marginales estimadas . . . . . 93
Ecuaciones de estimación generalizadas: Guardar . . . . . 95
Ecuaciones de estimación generalizadas: Exportar . . . . . 97
Funciones adicionales del comando GENLIN . . . . . 98

## 8 Modelos mixtos lineales generalizados 100

Obtención de un modelo mixto lineal generalizado . . . . . 102
Objetivo . . . . . 103
Efectos fijos . . . . . 106
Añadir un término personalizado . . . . . 107
Efectos aleatorios . . . . . 109
Bloque de efectos aleatorios . . . . . 110
Ponderación y desplazamiento . . . . . 112
Opciones de construcción . . . . . 113
Medias estimadas . . . . . 114
Guardar . . . . . 116
Vista del modelo . . . . . 117
Resumen del modelo . . . . . 117
Estructura de datos . . . . . 118
Pronóstico por observado . . . . . 118
Clasificación . . . . . 118
Efectos fijos . . . . . 118
Coeficientes fijos . . . . . 119
Covarianzas de efectos aleatorios . . . . . 120
Parámetros de covarianza . . . . . 120
Medias estimadas: Efectos significativos . . . . . 121
Medias estimadas: Efectos personalizados . . . . . 121

## 9 Análisis loglineal: Selección de modelo 123

Análisis loglineal: Definir rango . . . 124
Análisis loglineal: Modelo . . . 125
Construir términos . . . 125
Análisis loglineal: Opciones . . . 126
Funciones adicionales del comando HILOGLINEAR . . . 126

## 10 Análisis loglineal general 127

Análisis loglineal general: Modelo . . . 129
Construir términos . . . 129
Análisis loglineal general: Opciones . . . 130
Análisis loglineal general: Guardar . . . 131
Funciones adicionales del comando GENLOG . . . 131

## 11 Análisis loglineal logit 133

Análisis loglineal logit: Modelo . . . 135
Construir términos . . . 136
Análisis loglineal logit: Opciones . . . 136
Análisis loglineal logit: Guardar . . . 137
Funciones adicionales del comando GENLOG . . . 138

## 12 Tablas de mortalidad 139

Tablas de mortalidad: Definir evento para la variable de estado . . . 141
Tablas de mortalidad: Definir rango . . . 141
Tablas de mortalidad: Opciones . . . 142
Funciones adicionales del comando SURVIVAL . . . 142

## 13 Análisis de supervivencia de Kaplan-Meier 144

Kaplan-Meier: Definir evento para la variable de estado . . . 146
Kaplan-Meier: Comparar niveles de los factores . . . 146

Kaplan-Meier: Guardar variables nuevas . . . . . 147
Kaplan-Meier: Opciones . . . . . 148
Funciones adicionales del comando KM . . . . . 148

## 14 Análisis de regresión de Cox 150

Regresión de Cox: Definir variables categóricas . . . . . 152
Regresión de Cox: Gráficos . . . . . 153
Regresión de Cox: Guardar nuevas variables . . . . . 154
Regresión de Cox: Opciones . . . . . 155
Regresión de Cox: Definir evento para la variable de estado . . . . . 156
Funciones adicionales del comando COXREG . . . . . 156

## 15 Calcular covariable dependiente del tiempo 157

Para calcular una covariable dependiente del tiempo . . . . . 158
Regresión de Cox con covariables dependientes del tiempo: Funciones adicionales . . . . . 158

## Apéndices

## A Esquemas de codificación de variables categóricas 160

Desviación . . . . . 160
Simple . . . . . 161
Helmert . . . . . 162
Diferencia . . . . . 162
Polinómico . . . . . 163
Repetido . . . . . 163
Especial . . . . . 164
Indicador . . . . . 165

*B Estructuras de covarianza* *166*

*C Notices* *170*

*Índice* *172*

Capítulo

# 1

# Introducción a Estadísticas avanzadas

La opción Estadísticas avanzadas proporciona procedimientos que ofrecen opciones de modelado más avanzadas que las disponibles en el sistema Base.

- MLG Multivariado amplía el modelo lineal general que proporciona MLG Univariado al permitir varias variables dependientes. Una extensión adicional, GLM Medidas repetidas, permite las medidas repetidas de varias variables dependientes.
- Análisis de componentes de la varianza es una herramienta específica para descomponer la variabibilidad de una variable dependiente en componentes fijos y aleatorios.
- Los modelos mixtos lineales amplían el modelo lineal general de manera que los datos puedan presentar variabilidad correlacionada y no constante. El modelo lineal mixto proporciona, por tanto, la flexibilidad necesaria para modelar no sólo las medias sino también las varianzas y covarianzas de los datos.
- Los modelos lineales generalizados (GZLM) relajan el supuesto de normalidad del término de error y sólo requieren que la variable dependiente esté relacionada linealmente con los predictores mediante una transformación o función de enlace. Las ecuaciones de estimación generalizada (GEE) amplía GZLM para permitir medidas repetidas.
- El análisis loglineal general permite ajustar modelos a datos de recuento de clasificación cruzada y la selección del modelo del análisis loglineal puede ayudarle a elegir entre modelos.
- El análisis loglineal logit le permite ajustar modelos loglineales para analizar la relación existente entre una variable dependiente categórica y uno o más predictores categóricos.
- Puede realizar un análisis de supervivencia a través de Tablas de mortalidad para examinar la distribución de variables de tiempo de espera hasta un evento, posiblemente por niveles de una variable de factor; análisis de supervivencia de Kaplan-Meier para examinar la distribución de variables de tiempo de espera hasta un evento, posiblemente por niveles de una variable de factor o generar análisis separados por niveles de una variable de estratificación; y regresión de Cox para modelar el tiempo de espera hasta un determinado evento, basado en los valores de las covariables especificadas.

Capítulo
2

# Análisis MLG multivariante

El procedimiento MLG Multivariante proporciona un análisis de regresión y un análisis de varianza para variables dependientes múltiples por una o más covariables o variables de factor. Las variables de factor dividen la población en grupos. Utilizando este procedimiento de modelo lineal general, es posible contrastar hipótesis nulas sobre los efectos de las variables de factor sobre las medias de varias agrupaciones de una distribución conjunta de variables dependientes. Asimismo puede investigar las interacciones entre los factores y también los efectos individuales de los factores. Además, se pueden incluir los efectos de las covariables y las interacciones de covariables con los factores. Para el análisis de regresión, las variables (predictoras) independientes se especifican como covariables.

Se pueden contrastar tanto los modelos equilibrados como los no equilibrados. Se considera que un diseño está equilibrado si cada casilla del modelo contiene el mismo número de casos. En un modelo multivariado, las sumas de cuadrados debidas a los efectos del modelo y las sumas de cuadrados error se encuentran en forma de matriz en lugar de en la forma escalar del análisis univariado. Estas matrices se denominan matrices SCPC (sumas de cuadrados y productos cruzados). Si se especifica más de una variable dependiente, se proporciona el análisis multivariado de varianzas usando la traza de Pillai, la lambda de Wilks, la traza de Hotelling y el criterio de mayor raíz de Roy con el estadístico *F* aproximado, así como el análisis univariado de varianza para cada variable dependiente. Además de contratar hipótesis, MLG Multivariante genera estimaciones de los parámetros.

También se encuentran disponibles los contrastes *a priori* de uso más habitual para contrastar las hipótesis. Además, si una prueba *F* global ha mostrado cierta significación, pueden emplearse las pruebas post hoc para evaluar las diferencias entre las medias específicas. Las medias marginales estimadas ofrecen estimaciones de valores de las medias pronosticados para las casillas del modelo; los gráficos de perfil (gráficos de interacciones) de estas medias permiten observar fácilmente algunas de estas relaciones. Las pruebas de comparaciones múltiples post hoc se realizan por separado para cada variable dependiente.

En su archivo de datos puede guardar residuos, valores pronosticados, distancia de Cook y valores de influencia como variables nuevas para comprobar los supuestos. También se hallan disponibles una matriz SCPC residual, que es una matriz cuadrada de las sumas de cuadrados y los productos cruzados de los residuos; una matriz de covarianzas residual, que es la matriz SCPC residual dividida por los grados de libertad de los residuos; y la matriz de correlaciones residual, que es la forma tipificada de la matriz de covarianzas residual.

Ponderación MCP permite especificar una variable usada para aplicar a las observaciones una ponderación diferencial en un análisis de mínimos cuadrados ponderados (MCP), por ejemplo para compensar la distinta precisión de las medidas.

**Ejemplo.** Un fabricante de plásticos mide tres propiedades de la película de plástico: resistencia, brillo y opacidad. Se prueban dos tasas de extrusión y dos cantidades diferentes de aditivo y se miden las tres propiedades para cada combinación de tasa de extrusión y cantidad de aditivo. El fabricante deduce que la tasa de extrusión y la cantidad de aditivo producen individualmente resultados significativos, pero que la interacción de los dos factores no es significativa.

**Métodos.** Las sumas de cuadrados de Tipo I, Tipo II, Tipo III y Tipo IV pueden emplearse para evaluar las diferentes hipótesis. Tipo III es el valor por defecto.

**Estadísticos.** Las pruebas de rango post hoc y las comparaciones múltiples: Diferencia menos significativa (DMS), Bonferroni, Sidak, Scheffé, Múltiples *F* de Ryan-Einot-Gabriel-Welsch (R-E-G-W-F), Rango múltiple de Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls (S-N-K), Diferencia honestamente significativa de Tukey, *b* de Tukey, Duncan, GT2 de Hochberg, Gabriel, Pruebas *t* de Waller Duncan, Dunnett (unilateral y bilateral), T2 de Tamhane, T3 de Dunnett, Games-Howell y *C* de Dunnett. Estadísticos descriptivos: medias observadas, desviaciones típicas y recuentos de todas las variables dependientes en todas las casillas; la prueba de Levene sobre la homogeneidad de la varianza; la prueba *M* de Box sobre la homogeneidad de las matrices de covarianza de las variables dependientes; y la prueba de esfericidad de Bartlett.

**Diagramas.** Diagramas de dispersión por nivel, gráficos de residuos, gráficos de perfil (interacción).

**Datos.** Las variables dependientes deben ser cuantitativas. Los factores son categóricos y pueden tener valores numéricos o valores de cadena. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente.

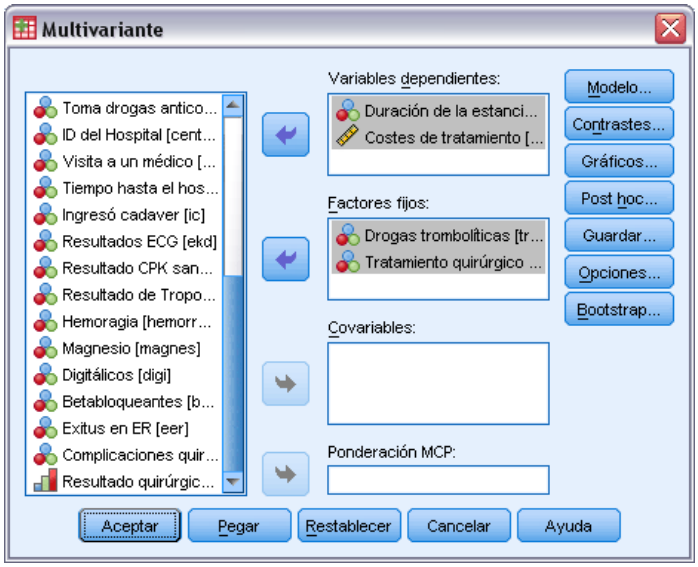
**Supuestos.** Para las variables dependientes, los datos son una muestra aleatoria de vectores de una población normal multivariada; en la población, las matrices de varianzas-covarianzas para todas las casillas son las mismas. El análisis de varianza es robusto a las desviaciones de la normalidad, aunque los datos deberán ser simétricos. Para comprobar los supuestos se pueden utilizar las pruebas de homogeneidad de varianzas (incluyendo la *M* de Box) y los gráficos de dispersión por nivel. También puede examinar los residuos y los gráficos de residuos.

**Procedimientos relacionados.** Utilice el procedimiento Explorar para examinar los datos antes de realizar un análisis de varianza. Para una variable dependiente única, utilice MLG Factorial General. Si ha medido las mismas variables dependientes en varias ocasiones para cada sujeto, utilice MLG Medidas repetidas.

### *Para obtener un análisis de varianza MLG multivariante*

- En los menús, seleccione:
  Analizar > Modelo lineal general > Multivariante...

Figura 2-1
*Cuadro de diálogo Multivariante*

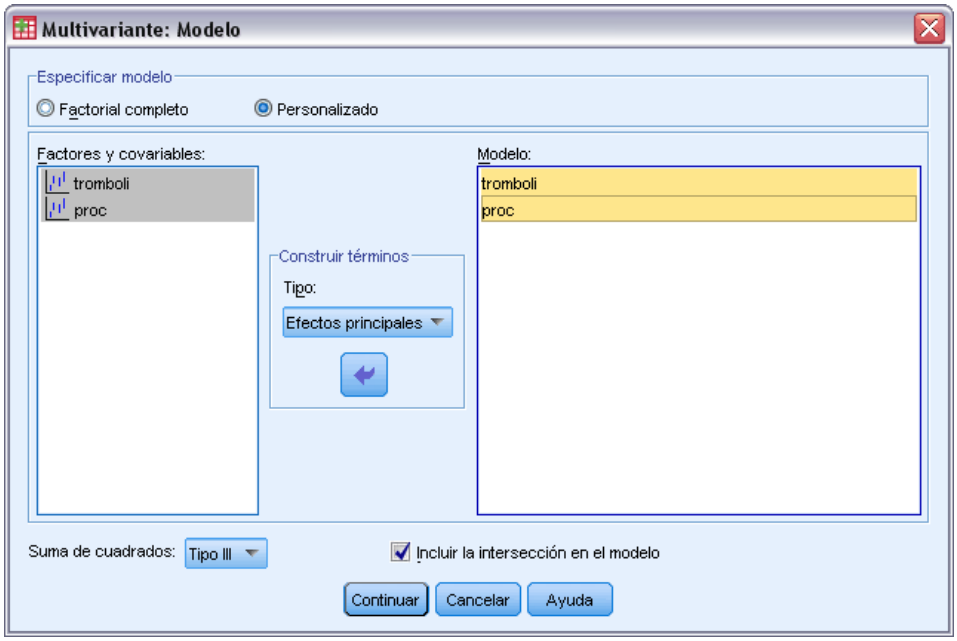



- Seleccione al menos dos variables dependientes.

  Si lo desea, puede especificar Factores fijos, Covariables y Ponderación MCP.

# *Modelo MLG multivariante*

Figura 2-2
*Cuadro de diálogo multivariante*



**Especificar modelo.** Un modelo factorial completo contiene todos los efectos principales del factor, todos los efectos principales de las covariables y todas las interacciones factor por factor. No contiene interacciones de covariable. Seleccione Personalizado para especificar sólo un subconjunto de interacciones o para especificar interacciones factor por covariable. Indique todos los términos que desee incluir en el modelo.

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

**Modelo.** El modelo depende de la naturaleza de los datos. Después de seleccionar Personalizado, puede elegir los efectos principales y las interacciones que sean de interés para el análisis.

**Suma de cuadrados** Determina el método para calcular las sumas de cuadrados. Para los modelos equilibrados y no equilibrados sin casillas perdidas, el método más utilizado para la suma de cuadrados es el Tipo III.

**Incluir la intersección en el modelo.** La intersección se incluye normalmente en el modelo. Si supone que los datos pasan por el origen, puede excluir la intersección.

## *Construir términos*

Para las covariables y los factores seleccionados:

**Interacción.** Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el método por defecto.

**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.

## *Suma de cuadrados*

Para el modelo, puede elegir un tipo de suma de cuadrados. El Tipo III es el más utilizado y es el tipo por defecto.

**Tipo I.** Este método también se conoce como el método de descomposición jerárquica de la suma de cuadrados. Cada término se corrige sólo respecto al término que le precede en el modelo. El método Tipo I para la obtención de sumas de cuadrados se utiliza normalmente para:

- Un modelo ANOVA equilibrado en el que se especifica cualquier efecto principal antes de cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se especifica antes de cualquier efecto de interacción de segundo orden, y así sucesivamente.

- Un modelo de regresión polinómica en el que se especifica cualquier término de orden inferior antes que cualquier término de orden superior.
- Un modelo puramente anidado en el que el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente. Esta forma de anidamiento solamente puede especificarse utilizando la sintaxis.

**Tipo II.** Este método calcula cada suma de cuadrados del modelo considerando sólo los efectos pertinentes. Un efecto pertinente es el que corresponde a todos los efectos que no contienen el que se está examinando. El método Tipo II para la obtención de sumas de cuadrados se utiliza normalmente para:

- Un modelo ANOVA equilibrado.
- Cualquier modelo que sólo tenga efectos de factor principal.
- Cualquier modelo de regresión.
- Un diseño puramente anidado (esta forma de anidamiento solamente puede especificarse utilizando la sintaxis).

**Tipo III.** Es el método por defecto. Este método calcula las sumas de cuadrados de un efecto del diseño como las sumas de cuadrados corregidas respecto a cualquier otro efecto que no lo contenga y ortogonales a cualquier efecto (si existe) que lo contenga. Las sumas de cuadrados de Tipo III tienen una gran ventaja por ser invariables respecto a las frecuencias de casilla, siempre que la forma general de estimabilidad permanezca constante. Así, este tipo de sumas de cuadrados se suele considerar de gran utilidad para un modelo no equilibrado sin casillas perdidas. En un diseño factorial sin casillas perdidas, este método equivale a la técnica de cuadrados ponderados de las medias de Yates. El método Tipo III para la obtención de sumas de cuadrados se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en los tipos I y II.
- Cualquier modelo equilibrado o desequilibrado sin casillas vacías.

**Tipo IV.** Este método está diseñado para una situación en la que hay casillas perdidas. Para cualquier efecto *F* en el diseño, si *F* no está contenida en cualquier otro efecto, entonces Tipo IV = Tipo III = Tipo II. Cuando *F* está contenida en otros efectos, el Tipo IV distribuye equitativamente los contrastes que se realizan entre los parámetros en *F* a todos los efectos de nivel superior. El método Tipo IV para la obtención de sumas de cuadrados se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en los tipos I y II.
- Cualquier modelo equilibrado o no equilibrado con casillas vacías.

# MLG Multivariante: Contrastes

Figura 2-3
*Cuadro de diálogo MLG Multivariante: Contrastes*


Los contrastes se utilizan para comprobar si los niveles de un efecto son significativamente diferentes unos de otros. Puede especificar un contraste para cada factor del modelo. Los contrastes representan las combinaciones lineales de los parámetros.

El contraste de hipótesis se basa en la hipótesis nula **LBM = 0**, donde **L** es la matriz de coeficientes de contraste, **M** es la matriz identidad (que tiene una dimensión igual al número de variables dependientes) y **B** es el vector de parámetros. Cuando se especifica un contraste, se crea una matriz **L** de modo que las columnas correspondientes al factor coincidan con el contraste. El resto de las columnas se corrigen para que la matriz **L** sea estimable.

Se ofrecen la prueba univariada que utiliza los estadísticos *F* y los intervalos de confianza simultáneos de tipo Bonferroni, basados en la distribución *t* de Student para las diferencias de contraste en todas las variables dependientes. También se ofrecen las pruebas multivariadas que utilizan los criterios de la traza de Pillai, la lambda de Wilks, la traza de Hotelling y la mayor raíz de Roy.

Los contrastes disponibles son de desviación, simples, de diferencias, de Helmert, repetidos y polinómicos. En los contrastes de desviación y los contrastes simples, es posible determinar que la categoría de referencia sea la primera o la última categoría.

## Tipos de contrastes

**Desviación.** Compara la media de cada nivel (excepto una categoría de referencia) con la media de todos los niveles (media global). Los niveles del factor pueden colocarse en cualquier orden.

**Simple.** Compara la media de cada nivel con la media de un nivel especificado. Este tipo de contraste resulta útil cuando existe un grupo de control. Puede seleccionar la primera o la última categoría como referencia.

**Diferencia.** Compara la media de cada nivel (excepto el primero) con la media de los niveles anteriores (a veces también se denominan contrastes de Helmert inversos). (a veces también se denominan contrastes de Helmert inversos).

**Helmert.** Compara la media de cada nivel del factor (excepto el último) con la media de los niveles siguientes.

**Repetidas.** Compara la media de cada nivel (excepto el último) con la media del nivel siguiente.

**Polinómico.** Compara el efecto lineal, cuadrático, cúbico, etc. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, y así sucesivamente. Estos contrastes se utilizan a menudo para estimar las tendencias polinómicas.

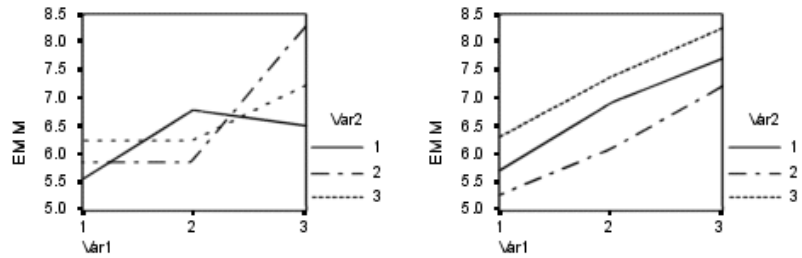
## Gráficos de perfil de MLG multivariante

Figura 2-4
*Cuadro de diálogo Multivariante: Gráficos de perfil*


Los gráficos de perfil (gráficos de interacción) sirven para comparar las medias marginales en el modelo. Un gráfico de perfil es un gráfico de líneas en el que cada punto indica la media marginal estimada de una variable dependiente (corregida respecto a las covariables) en un nivel de un factor. Los niveles de un segundo factor se pueden utilizar para generar líneas diferentes. Cada nivel en un tercer factor se puede utilizar para crear un gráfico diferente. Todos los factores están disponibles para los gráficos. Los gráficos de perfil se crean para cada variable dependiente.

Un gráfico de perfil de un factor muestra si las medias marginales estimadas aumentan o disminuyen a través de los niveles. Para dos o más factores, las líneas paralelas indican que no existe interacción entre los factores, lo que significa que puede investigar los niveles de un único factor. Las líneas no paralelas indican una interacción.

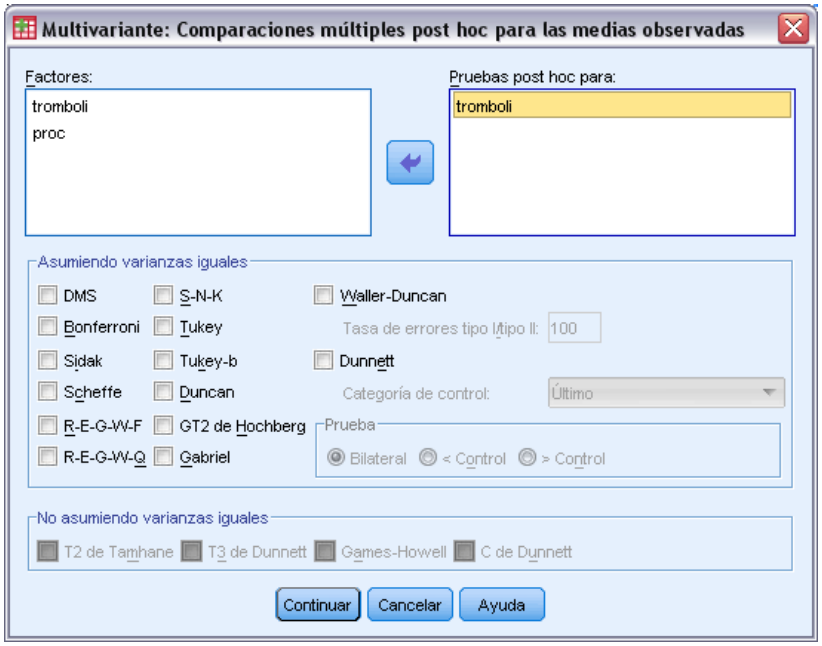
Figura 2-5
*Gráfico no paralelo (izquierda) y gráfico paralelo (derecha)*



Después de especificar un gráfico mediante la selección de los factores del eje horizontal y, de manera opcional, los factores para distintas líneas y gráficos, el gráfico deberá añadirse a la lista de gráficos.

## MLG multivariante: Comparaciones post hoc

Figura 2-6
*Cuadro de diálogo Multivariante: Comparaciones múltiples post hoc para las medias observadas*




**Pruebas de comparaciones múltiples post hoc** Una vez que se ha determinado que existen diferencias entre las medias, las pruebas de rango post hoc y las comparaciones múltiples por parejas permiten determinar qué medias difieren. Las comparaciones se realizan sobre valores sin corregir. Las pruebas post hoc se realizan por separado para cada variable dependiente.

Las pruebas de diferencia honestamente significativa de Tukey y de Bonferroni son pruebas de comparación múltiple muy utilizadas. La **prueba de Bonferroni**, basada en el estadístico *t* de Student, corrige el nivel de significación observado por el hecho de que se realizan comparaciones múltiples. La **prueba t de Sidak** también corrige el nivel de significación y da lugar a límites más estrechos que los de Bonferroni. La **prueba de diferencia honestamente significativa de Tukey** utiliza el estadístico del rango estudentizado para realizar todas las comparaciones por pares entre los grupos y establece la tasa de error por experimento como la tasa de error para el conjunto de todas las comparaciones por pares. Cuando se contrasta un gran número de pares de medias, la prueba de la diferencia honestamente significativa de Tukey es más potente que la prueba de Bonferroni. Para un número reducido de pares, Bonferroni es más potente.

**GT2 de Hochberg** es similar a la prueba de la diferencia honestamente significativa de Tukey, pero se utiliza el módulo máximo estudentizado. La prueba de Tukey suele ser más potente. La **prueba de comparación por parejas de Gabriel** también utiliza el módulo máximo estudentizado y es generalmente más potente que la GT2 de Hochberg cuando los tamaños de las

casillas son desiguales. La prueba de Gabriel se puede convertir en liberal cuando los tamaños de las casillas varían mucho.

La **prueba t de comparación múltiple por parejas de Dunnett** compara un conjunto de tratamientos con una media de control simple. La última categoría es la categoría de control por defecto. Si lo desea, puede seleccionar la primera categoría. Asimismo, puede elegir una prueba unilateral o bilateral. Para comprobar que la media de cualquier nivel del factor (excepto la categoría de control) no es igual a la de la categoría de control, utilice una prueba bilateral. Para contrastar si la media en cualquier nivel del factor es menor que la de la categoría de control, seleccione < Control. Asimismo, para contrastar si la media en cualquier nivel del factor es mayor que la de la categoría de control, seleccione > Control.

Ryan, Einot, Gabriel y Welsch (R-E-G-W) desarrollaron dos pruebas de rangos múltiples por pasos. Los procedimientos múltiples por pasos (por tamaño de las distancias) contrastan en primer lugar si todas las medias son iguales. Si no son iguales, se contrasta la igualdad en los subconjuntos de medias. **R-E-G-W F** se basa en una prueba *F* y **R-E-G-W Q** se basa en un rango estudentizado. Estas pruebas son más potentes que la prueba de rangos múltiples de Duncan y Student-Newman-Keuls (que también son procedimientos múltiples por pasos), pero no se recomiendan para tamaños de casillas desiguales.

Cuando las varianzas son desiguales, utilice **T2 de Tamhane** (prueba conservadora de comparación por parejas basada en una prueba *t*), **T3 de Dunnett** (prueba de comparación por parejas basada en el módulo máximo estudentizado), **prueba de comparación por parejasGames-Howell** (a veces liberal), o **C de Dunnett** (prueba de comparación por parejas basada en el rango estudentizado).

La **prueba de rango múltiple de Duncan**, Student-Newman-Keuls (**S-N-K**) y **b de Tukey** son pruebas de rango que asignan rangos a medias de grupo y calculan un valor de rango. Estas pruebas no se utilizan con la misma frecuencia que las pruebas anteriormente mencionadas.

La **prueba t de Waller-Duncan** utiliza la aproximación bayesiana. Esta prueba de rango emplea la media armónica del tamaño muestral cuando los tamaños muestrales no son iguales.

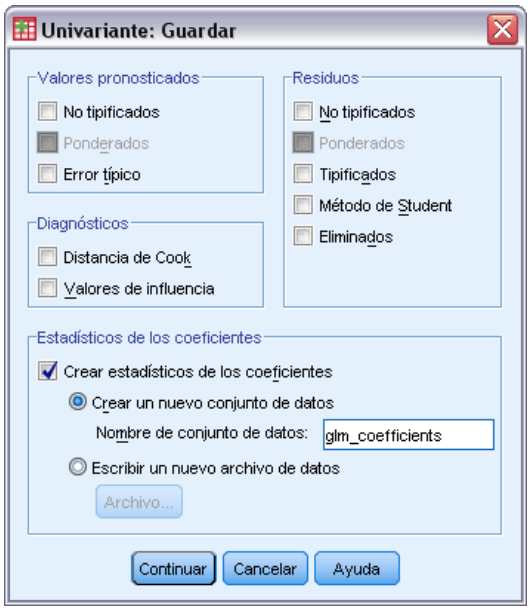
El nivel de significación de la prueba de **Scheffé** está diseñado para permitir todas las combinaciones lineales posibles de las medias de grupo que se van a contrastar, no sólo las comparaciones por parejas disponibles en esta función. El resultado es que la prueba de Scheffé es normalmente más conservadora que otras pruebas, lo que significa que se precisa una mayor diferencia entre las medias para la significación.

La prueba de comparación múltiple por parejas de la diferencia menos significativa (**DMS**) es equivalente a varias pruebas *t* individuales entre todos los pares de grupos. La desventaja de esta prueba es que no se realiza ningún intento de corregir el nivel crítico para realizar las comparaciones múltiples.

**Pruebas mostradas.** Se proporcionan comparaciones por parejas para DMS, Sidak, Bonferroni, Games-Howell, T2 y T3 de Tamhane, *C* de Dunnett y T3 de Dunnett. También se facilitan subconjuntos homogéneos para S-N-K, *b* de Tukey, Duncan, R-E-G-W *F*, R-E-G-W *Q* y Waller. La prueba de la diferencia honestamente significativa de Tukey, la GT2 de Hochberg, la prueba de Gabriel y la prueba de Scheffé son pruebas de comparaciones múltiples y pruebas de rango.

# MLG: Guardar

Figura 2-7
*Cuadro de diálogo Guardar*




Es posible guardar los valores pronosticados por el modelo, los residuos y las medidas relacionadas como variables nuevas en el Editor de datos. Muchas de estas variables se pueden utilizar para examinar supuestos sobre los datos. Si desea almacenar los valores para utilizarlos en otra sesión de IBM® SPSS® Statistics, guárdelos en el archivo de datos actual.

**Valores pronosticados.** Son los valores que predice el modelo para cada caso.

- **No tipificados.** Valor pronosticado por el modelo para la variable dependiente.
- **Ponderados.** Valores pronosticados no tipificados ponderados. Sólo están disponibles si se seleccionó previamente una variable de ponderación MCP.
- **Error típico.** Estimación de la desviación típica del valor promedio de la variable dependiente para los casos que tengan los mismos valores en las variables independientes.

**Diagnósticos.** Son medidas para identificar casos con combinaciones poco usuales de valores para los casos y las variables independientes que puedan tener un gran impacto en el modelo.

- **Distancia 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. Una Distancia de Cook grande indica que la exclusión de ese caso del cálculo de los estadísticos de regresión hará variar substancialmente los coeficientes.
- **Valores de influencia.** Valores de influencia no centrados. La influencia relativa de una observación en el ajuste del modelo.

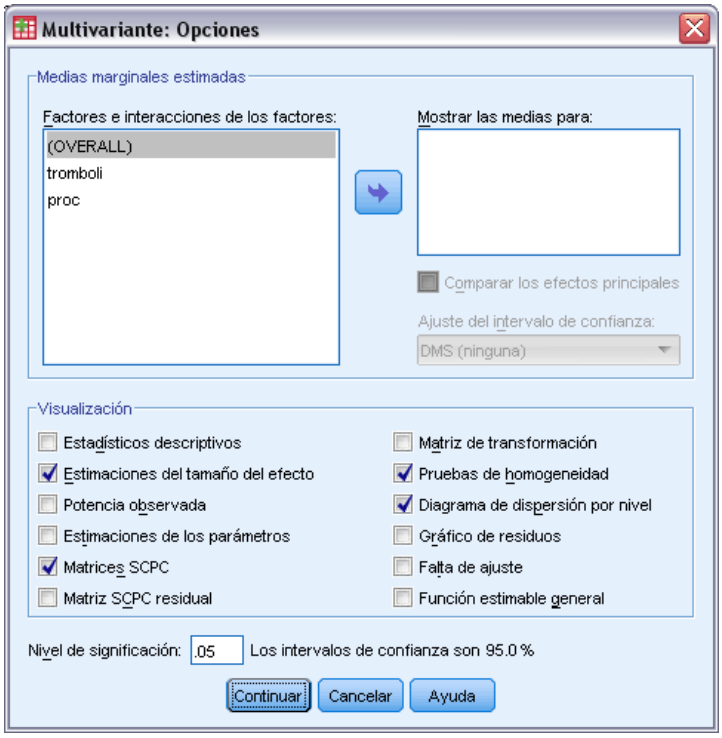
**Residuos.** Un residuo no tipificado es el valor real de la variable dependiente menos el valor pronosticado por el modelo. También se encuentran disponibles residuos eliminados, estudentizados y tipificados. Si ha seleccionado una variable MCP, contará además con residuos no tipificados ponderados.

- **No tipificados.** Diferencia entre un valor observado y el valor pronosticado por el modelo.
- **Ponderados.** Residuos no tipificados ponderados. Sólo están disponibles si se seleccionó previamente una variable de ponderación MCP.
- **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.
- **Método de Student.** Residuo dividido por una estimación de su desviación típica que varía de caso en caso, dependiendo de la distancia de los valores de cada caso en las variables independientes respecto a las medias en las variables independientes.
- **Eliminados.** Residuo para un caso cuando éste se excluye del cálculo de los coeficientes de la regresión. Es igual a la diferencia entre el valor de la variable dependiente y el valor pronosticado corregido.

**Estadísticos de los coeficientes.** Escribe una matriz varianza-covarianza de las estimaciones de los parámetros del modelo en un nuevo conjunto de datos de la sesión actual o un archivo de datos externo de SPSS Statistics. Asimismo, para cada variable dependiente habrá una fila de estimaciones de los parámetros, una fila de valores de significación para los estadísticos *t* correspondientes a las estimaciones de los parámetros y una fila de grados de libertad de los residuos. En un modelo multivariante, existen filas similares para cada variable dependiente. Si lo desea, puede usar este archivo matricial en otros procedimientos que lean archivos matriciales.

# *MLG Multivariante: Opciones*

Figura 2-8
*Cuadro de diálogo MLG Multivariante: Opciones*



Este cuadro de diálogo contiene estadísticos opcionales. Los estadísticos se calculan utilizando un modelo de efectos fijos.

**Medias marginales estimadas.** Seleccione los factores e interacciones para los que desee obtener estimaciones de las medias marginales de la población en las casillas. Estas medias se corrigen respecto a las covariables, si las hay. Las interacciones sólo están disponibles si se ha especificado un modelo personalizado.

- **Comparar los efectos principales.** Proporciona comparaciones por parejas no corregidas entre las medias marginales estimadas para cualquier efecto principal del modelo, tanto para los factores inter-sujetos como para los intra-sujetos. Este elemento sólo se encuentra disponible si los efectos principales están seleccionados en la lista Mostrar las medias para.
- **Ajuste del intervalo de confianza.** Seleccione un ajuste de diferencia menor significativa (DMS), Bonferroni o Sidak para los intervalos de confianza y la significación. Este elemento sólo estará disponible si se selecciona Comparar los efectos principales.

**Mostrar.** Seleccione Estadísticos descriptivos para generar medias observadas, desviaciones típicas y frecuencias para cada variable dependiente en todas las casillas. La opción Estimaciones del tamaño del efecto ofrece un valor parcial de eta-cuadrado para cada efecto y cada estimación de parámetros. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor. Seleccione Potencia observada para obtener la potencia de la prueba cuando la hipótesis alternativa se ha establecido basándose en el valor observado. Seleccione Estimaciones de los parámetros para generar las estimaciones de los parámetros, los errores típicos, las pruebas *t*, los intervalos de confianza y la potencia observada para cada prueba. Se pueden mostrar Matrices SCPC de error y de hipótesis y la Matriz SCPC residual más la prueba de esfericidad de Bartlett de la matriz de covarianzas residual.

Las pruebas de homogeneidad producen la prueba de homogeneidad de varianzas de Levene para cada variable dependiente en todas las combinaciones de nivel de los factores inter-sujetos sólo para factores inter-sujetos. Asimismo, las pruebas de homogeneidad incluyen la prueba *M* de Box sobre la homogeneidad de las matrices de covarianza de las variables dependientes a lo largo de todas las combinaciones de niveles de los factores inter-sujetos. Las opciones de diagramas de dispersión por nivel y gráfico de los residuos son útiles para comprobar los supuestos sobre los datos. Estos elementos no estarán activado si no hay factores. Seleccione Gráficos de los residuos para generar un gráfico de los residuos observados respecto a los pronosticados respecto a los tipificados para cada variable dependiente. Estos gráficos son útiles para investigar el supuesto de varianzas iguales. Seleccione la Prueba de falta de ajuste para comprobar si el modelo puede describir de forma adecuada la relación entre la variable dependiente y las variables independientes. La función estimable general permite construir pruebas de hipótesis personales basadas en la función estimable general. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función estimable general.

**Nivel de significación.** Puede que le interese corregir el nivel de significación usado en las pruebas post hoc y el nivel de confianza empleado para construir intervalos de confianza. El valor especificado también se utiliza para calcular la potencia observada para la prueba. Si especifica un nivel de significación, el cuadro de diálogo mostrará el nivel asociado de los intervalos de confianza.

# *Funciones adicionales del comando GLM*

Estas funciones se pueden aplicar a los análisis univariados, multivariados o de medidas repetidas. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar efectos anidados en el diseño (utilizando el subcomando `DESIGN`).
- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando `TEST`).
- Especificar contrastes múltiples (utilizando el subcomando `CONTRAST`).
- Incluir los valores perdidos definidos por el usuario (utilizando el subcomando `MISSING`).
- Especificar criterios EPS (mediante el subcomando `CRITERIA`).
- Construir una matriz**L**, una matriz **M** o una matriz **K** (utilizando los subcomandos `LMATRIX`, `MMATRIX` o `KMATRIX`).
- Especificar una categoría de referencia intermedia (utilizando el subcomando `CONTRAST` para los contrastes de desviación o simples).
- Especificar la métrica para los contrastes polinómicos (utilizando el subcomando `CONTRAST`).
- Especificar términos de error para las comparaciones post hoc (utilizando el subcomando `POSTHOC`).
- Calcular medias marginales estimadas para cualquier factor o interacción entre los factores en la lista de factores (utilizando el subcomando `EMMEANS`).
- Especificar nombres para las variables temporales (utilizando el subcomando `SAVE`).
- Construir un archivo de datos matricial de correlaciones (utilizando el subcomando `OUTFILE`).
- Construir un archivo de datos matricial que contenga estadísticos de la tabla de ANOVA inter-sujetos (utilizando el subcomando `OUTFILE`).
- Guardar la matriz del diseño en un nuevo archivo de datos (utilizando el subcomando `OUTFILE`).

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

Capítulo
3

# MLG Medidas repetidas

El procedimiento MLG Medidas repetidas proporciona un análisis de varianza cuando se toma la misma medida varias veces a cada sujeto o caso. Si se especifican factores inter-sujetos, éstos dividen la población en grupos. Utilizando este procedimiento de modelo lineal general, puede contrastar hipótesis nulas sobre los efectos tanto de los factores inter-sujetos como de los factores intra-sujetos. Asimismo puede investigar las interacciones entre los factores y también los efectos individuales de los factores. También se pueden incluir los efectos de covariables constantes y de las interacciones de las covariables con los factores inter-sujetos.

En un diseño doblemente multivariado de medidas repetidas, las variables dependientes representan medidas de más de una variable para los diferentes niveles de los factores intra-sujetos. Por ejemplo, se pueden haber medido el pulso y la respiración de cada sujeto en tres momentos diferentes.

El procedimiento MLG Medidas repetidas ofrece análisis univariados y multivariados para datos de medidas repetidas. Se pueden contrastar tanto los modelos equilibrados como los no equilibrados. Se considera que un diseño está equilibrado si cada casilla del modelo contiene el mismo número de casos. En un modelo multivariado, las sumas de cuadrados debidas a los efectos del modelo y las sumas de cuadrados error se encuentran en forma de matriz en lugar de en la forma escalar del análisis univariado. Estas matrices se denominan matrices SCPC (sumas de cuadrados y productos cruzados). Además de contrastar las hipótesis, MLG Medidas repetidas genera estimaciones de los parámetros.

Se encuentran disponibles los contrastes *a priori* utilizados habitualmente para elaborar hipótesis que contrastan los factores inter-sujetos. Además, si una prueba *F* global ha mostrado cierta significación, pueden emplearse las pruebas post hoc para evaluar las diferencias entre las medias específicas. Las medias marginales estimadas ofrecen estimaciones de valores de las medias pronosticados para las casillas del modelo; los gráficos de perfil (gráficos de interacciones) de estas medias permiten observar fácilmente algunas de estas relaciones.

En su archivo de datos puede guardar residuos, valores pronosticados, distancia de Cook y valores de influencia como variables nuevas para comprobar los supuestos. También se hallan disponibles una matriz SCPC residual, que es una matriz cuadrada de las sumas de cuadrados y los productos cruzados de los residuos; una matriz de covarianzas residual, que es la matriz SCPC residual dividida por los grados de libertad de los residuos; y la matriz de correlaciones residual, que es la forma tipificada de la matriz de covarianzas residual.

Ponderación MCP permite especificar una variable usada para aplicar a las observaciones una ponderación diferencial en un análisis de mínimos cuadrados ponderados (MCP), por ejemplo para compensar la distinta precisión de las medidas.

**Ejemplo.** Se asignan doce estudiantes a un grupo de alta o de baja ansiedad basándose en las puntuaciones obtenidas en una prueba de nivel de ansiedad. El nivel de ansiedad es un factor inter-sujetos porque divide a los sujetos en grupos. A cada estudiante se le dan cuatro ensayos para

una determinada tarea de aprendizaje y se registra el número de errores por ensayo. Los errores de cada ensayo se registran en variables distintas y se define un factor intra-sujetos (ensayo) con cuatro niveles para cada uno de los cuatro ensayos. Se descubre que el efecto de los ensayos es significativo, mientras que la interacción ensayo-ansiedad no es significativa.

**Métodos.** Las sumas de cuadrados de Tipo I, Tipo II, Tipo III y Tipo IV pueden emplearse para evaluar las diferentes hipótesis. Tipo III es el valor por defecto.

**Estadísticos.** Pruebas de rango post hoc y comparaciones múltiples (para los factores inter-sujetos): Diferencia menos significativa (DMS), Bonferroni, Sidak, Scheffé, Múltiples *F* de Ryan-Einot-Gabriel-Welsch (R-E-G-W-F), Rango múltiple de Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls (S-N-K), Diferencia honestamente significativa de Tukey, *b* de Tukey, Duncan, GT2 de Hochberg, Gabriel, Pruebas *t* de Waller Duncan, Dunnett (unilateral y bilateral), T2 de Tamhane, T3 de Dunnett, Games-Howell y *C* de Dunnett. Estadísticos descriptivos: medias observadas, desviaciones típicas y recuentos de todas las variables dependientes en todas las casillas; la prueba de Levene sobre la homogeneidad de la varianza; la *M* de Box; y la prueba de esfericidad de Mauchly.

**Diagramas.** Diagramas de dispersión por nivel, gráficos de residuos, gráficos de perfil (interacción).

**Datos.** Las variables dependientes deben ser cuantitativas. Los factores inter-sujetos dividen la muestra en subgrupos discretos, como hombre y mujer. Estos factores son categóricos y pueden tener valores numéricos o valores de cadena. Los factores intra-sujetos se definen en el cuadro de diálogo MLG Medidas repetidas: Definir factores. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente. Para un análisis de medidas repetidas, las covariables deberán permanecer constantes en cada nivel de la variable intra-sujetos.

El archivo de datos debe contener un conjunto de variables para cada grupo de medidas tomadas a los sujetos. El conjunto tiene una variable para cada repetición de la medida dentro del grupo. Se define un factor intra-sujetos para el grupo con el número de niveles igual al número de repeticiones. Por ejemplo, se podrían tomar medidas del peso en días diferentes. Si las medidas de esa misma propiedad se han tomado durante cinco días, el factor intra-sujetos podría especificarse como *día* con cinco niveles.

Para múltiples factores intra-sujetos, el número de medidas de cada sujeto es igual al producto del número de niveles de cada factor. Por ejemplo, si las mediciones se tomaran en tres momentos diferentes del día durante cuatro días, el número total de medidas sería 12 para cada sujeto. Los factores intra-sujetos podrían especificarse como *día(4)* y *mediciones(3)*.

**Supuestos.** Un análisis de medidas repetidas se puede enfocar de dos formas: univariado y multivariado.

El enfoque univariado (también conocido como el método de modelo mixto o split-plot) considera las variables dependientes como respuestas a los niveles de los factores intra-sujetos. Las medidas en un sujeto deben ser una muestra de una distribución normal multivariada y las matrices de varianzas-covarianzas son las mismas en todas las casillas formadas por los efectos inter-sujetos. Se realizan ciertos supuestos sobre la matriz de varianzas-covarianzas de las variables dependientes. La validez del estadístico *F* utilizado en el enfoque univariado puede garantizarse si la matriz de varianzas-covarianzas es de forma circular (Huynh y Mandeville, 1979).

Para contrastar este supuesto se puede utilizar la prueba de esfericidad de Mauchly, que realiza una prueba de esfericidad sobre la matriz de varianzas-covarianzas de la variable dependiente transformada y ortonormalizada. La prueba de Mauchly aparece automáticamente en el análisis de medidas repetidas. En las muestras de tamaño reducido, esta prueba no resulta muy potente. En las de gran tamaño, la prueba puede ser significativa incluso si es pequeño el impacto de la desviación en los resultados. Si la significación de la prueba es grande, se puede asumir la hipótesis de esfericidad. Sin embargo, si la significación es pequeña y parece que se ha violado el supuesto de esfericidad, se puede realizar una corrección en los grados de libertad del numerador y del denominador para validar el estadístico *F* univariado. Se encuentran disponibles tres estimaciones para dicha corrección, denominada **épsilon**, en el procedimiento MLG Medidas repetidas. Los grados de libertad tanto del numerador como del denominador deben multiplicarse por épsilon y la significación del cociente *F* debe evaluarse con los nuevos grados de libertad.

El enfoque multivariado considera que las medidas de un sujeto son una muestra de una distribución normal multivariada y las matrices de varianzas-covarianzas son las mismas en todas las casillas formadas por los efectos inter-sujetos. Para contrastar si las matrices de varianzas-covarianzas de todas las casillas son las mismas, se puede utilizar la prueba *M* de Box.

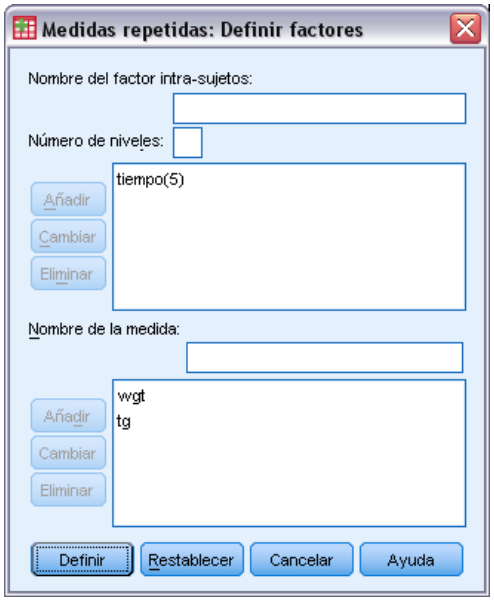
**Procedimientos relacionados.** Utilice el procedimiento Explorar para examinar los datos antes de realizar un análisis de varianza. Si *no* existen medidas repetidas para cada sujeto, utilice MLG Univariante o MLG Multivariante. Si sólo existen dos medidas para cada sujeto (por ejemplo, un pre-test y un post-test) y no hay factores inter-sujetos, puede utilizar el procedimiento Prueba T para muestras relacionadas.

### *Obtención de MLG Medidas repetidas*

▶ Elija en los menús:

Analizar > Modelo lineal general > Medidas repetidas...

Figura 3-1
*Cuadro de diálogo MLG Medidas repetidas: Definir factores*



- Escriba un nombre para el factor intra-sujetos y su número de niveles.
- Pulse en Añadir.
- Repita estos pasos para cada factor intra-sujetos.

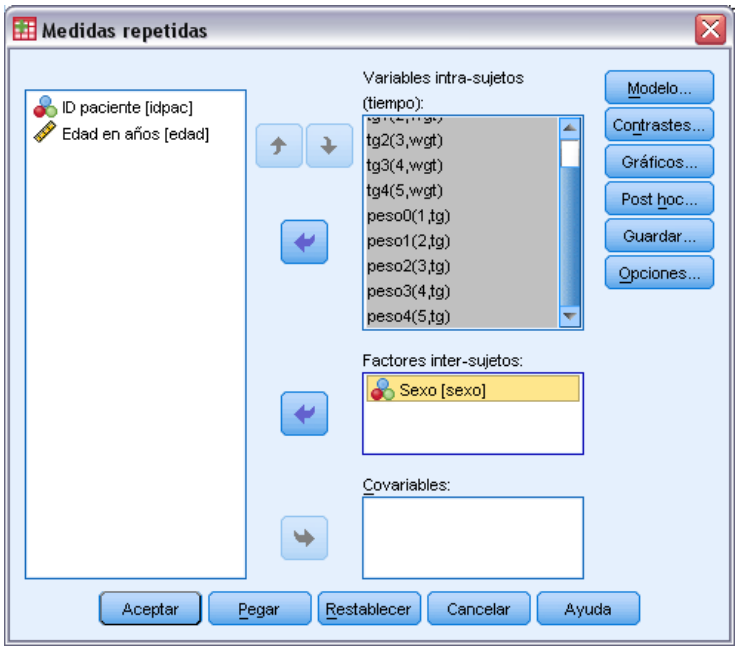
Para definir factores de medidas en un diseño doblemente multivariado de medidas repetidas:

- Escriba el nombre de la medida.
- Pulse en Añadir.

Después de definir todos los factores y las medidas:

- Pulse en Definir.

Figura 3-2
*Cuadro de diálogo MLG Medidas repetidas*




- Seleccione en la lista una variable dependiente que corresponda a cada combinación de factores intra-sujetos (y, de forma opcional, medidas).

Para cambiar las posiciones de las variables, utilice las flechas arriba y abajo.

Para realizar cambios en los factores intra-sujetos, puede volver a abrir el cuadro de diálogo MLG Medidas repetidas: Definir factores sin cerrar el cuadro de diálogo principal. Si lo desea, puede especificar covariables y factores inter-sujetos.

## MLG Medidas repetidas: Definir factores

MLG Medidas repetidas analiza grupos de variables dependientes relacionadas que representan diferentes medidas del mismo atributo. Este cuadro de diálogo permite definir uno o varios factores intra-sujetos para utilizarlos en MLG Medidas repetidas. Consulte Figura 3-1 el p. 17.

Tenga en cuenta que el orden en el que se especifiquen los factores intra-sujetos es importante. Cada factor constituye un nivel dentro del factor precedente.

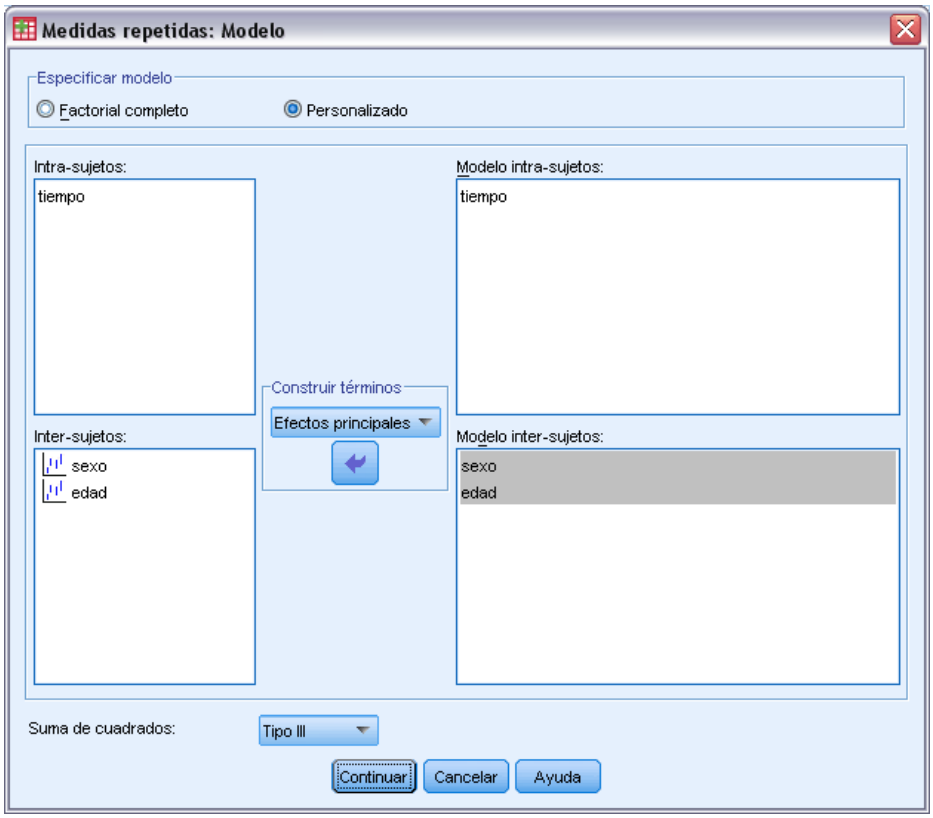
Para utilizar Medidas repetidas, deberá definir los datos correctamente. Los factores intra-sujetos deben definirse en este cuadro de diálogo. Observe que estos factores no son las variables existentes en sus datos, sino los factores que deberá definir aquí.

**Ejemplo.** En un estudio sobre la pérdida de peso, suponga que se mide cada semana el peso de varias personas durante cinco semanas. En el archivo de datos, cada persona es un sujeto o caso. Los pesos de las distintas semanas se registran en las variables *peso1*, *peso2*, etc. El sexo de cada persona se registra en otra variable. Los pesos, medidos repetidamente para cada sujeto, se pueden agrupar definiendo un factor intra-sujetos. Este factor podría denominarse *semana*, definido con cinco niveles. En el cuadro de diálogo principal, las variables *peso1*, ..., *peso5* se utilizan para asignar los cinco niveles de *semana*. La variable del archivo de datos que agrupa a hombres y mujeres (*sexo*) puede especificarse como un factor inter-sujetos, para estudiar las diferencias entre hombres y mujeres.

**Medidas.** Si los sujetos se comparan en más de una medida cada vez, defina las medidas. Por ejemplo, se podría medir el ritmo de la respiración y el pulso para cada sujeto todos los días durante una semana. El nombre de las medidas no existen como un nombre de variables en el propio archivo de datos sino se define aquí. Un modelo con más de una medida a veces se denomina modelo doblemente multivariado de medidas repetidas.

# MLG Medidas repetidas: Modelo

Figura 3-3
*Cuadro de diálogo MLG Medidas repetidas: Modelo*




**Especificar modelo.** Un modelo factorial completo contiene todos los efectos principales del factor, todos los efectos principales de las covariables y todas las interacciones factor por factor. No contiene interacciones de covariable. Seleccione Personalizado para especificar sólo un subconjunto de interacciones o para especificar interacciones factor por covariable. Indique todos los términos que desee incluir en el modelo.

**Inter-sujetos.** Muestra una lista de los factores inter-sujetos y las covariables.

**Modelo.** El modelo depende de la naturaleza de los datos. Tras elegir Personalizado, puede seleccionar los efectos y las interacciones intra-sujetos y los efectos y las interacciones inter-sujetos que sean de interés para el análisis.

**Suma de cuadrados** Determina el método de cálculo de las sumas de cuadrados para el modelo inter-sujetos. Para los modelos inter-sujetos equilibrados y no equilibrados sin casillas perdidas, el método más utilizado para la suma de cuadrados es el Tipo III.

## *Construir términos*

Para las covariables y los factores seleccionados:

**Interacción.** Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el método por defecto.

**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.

## *Suma de cuadrados*

Para el modelo, puede elegir un tipo de suma de cuadrados. El Tipo III es el más utilizado y es el tipo por defecto.

**Tipo I.** Este método también se conoce como el método de descomposición jerárquica de la suma de cuadrados. Cada término se corrige sólo respecto al término que le precede en el modelo. El método Tipo I para la obtención de sumas de cuadrados se utiliza normalmente para:

- Un modelo ANOVA equilibrado en el que se especifica cualquier efecto principal antes de cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se especifica antes de cualquier efecto de interacción de segundo orden, y así sucesivamente.
- Un modelo de regresión polinómica en el que se especifica cualquier término de orden inferior antes que cualquier término de orden superior.
- Un modelo puramente anidado en el que el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente. Esta forma de anidamiento solamente puede especificarse utilizando la sintaxis.

**Tipo II.** Este método calcula cada suma de cuadrados del modelo considerando sólo los efectos pertinentes. Un efecto pertinente es el que corresponde a todos los efectos que no contienen el que se está examinando. El método Tipo II para la obtención de sumas de cuadrados se utiliza normalmente para:

- Un modelo ANOVA equilibrado.
- Cualquier modelo que sólo tenga efectos de factor principal.
- Cualquier modelo de regresión.
- Un diseño puramente anidado (esta forma de anidamiento solamente puede especificarse utilizando la sintaxis).

**Tipo III.** Es el método por defecto. Este método calcula las sumas de cuadrados de un efecto del diseño como las sumas de cuadrados corregidas respecto a cualquier otro efecto que no lo contenga y ortogonales a cualquier efecto (si existe) que lo contenga. Las sumas de cuadrados de Tipo III tienen una gran ventaja por ser invariables respecto a las frecuencias de casilla, siempre

que la forma general de estimabilidad permanezca constante. Así, este tipo de sumas de cuadrados se suele considerar de gran utilidad para un modelo no equilibrado sin casillas perdidas. En un diseño factorial sin casillas perdidas, este método equivale a la técnica de cuadrados ponderados de las medias de Yates. El método Tipo III para la obtención de sumas de cuadrados se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en los tipos I y II.
- Cualquier modelo equilibrado o desequilibrado sin casillas vacías.

**Tipo IV.** Este método está diseñado para una situación en la que hay casillas perdidas. Para cualquier efecto *F* en el diseño, si *F* no está contenida en cualquier otro efecto, entonces Tipo IV = Tipo III = Tipo II. Cuando *F* está contenida en otros efectos, el Tipo IV distribuye equitativamente los contrastes que se realizan entre los parámetros en *F* a todos los efectos de nivel superior. El método Tipo IV para la obtención de sumas de cuadrados se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en los tipos I y II.
- Cualquier modelo equilibrado o no equilibrado con casillas vacías.

## MLG Medidas repetidas: Contrastes

Figura 3-4
*Cuadro de diálogo MLG Medidas repetidas: Contrastes*


Los contrastes se utilizan para contrastar las diferencias entre los niveles de un factor inter-sujetos. Puede especificar un contraste para cada factor inter-sujetos del modelo. Los contrastes representan las combinaciones lineales de los parámetros.

El contraste de hipótesis se basa en la hipótesis nula **LBM** = 0, donde **L** es la matriz de coeficientes de contraste, **B** es el vector de parámetros y **M** es la matriz promedio que corresponde a la transformación promedio para la variable dependiente. Puede mostrar esta matriz de transformación seleccionando la opción Matriz de transformación en el cuadro de diálogo Medidas repetidas: Opciones. Por ejemplo, si existen cuatro variables dependientes, un factor intra-sujetos de cuatro niveles y se utilizan contrastes polinómicos (valor por defecto) para los factores intra-sujetos, la matriz **M** será (0,5 0,5 0,5 0,5)'. Cuando se especifica un contraste, se crea una matriz **L** de modo que las columnas correspondientes al factor inter-sujetos coincidan con el contraste. El resto de las columnas se corrigen para que la matriz **L** sea estimable.

Los contrastes disponibles son de desviación, simples, de diferencias, de Helmert, repetidos y polinómicos. En los contrastes de desviación y los contrastes simples, es posible determinar que la categoría de referencia sea la primera o la última categoría.

Deberá seleccionar un contraste que no sea Ninguno para factores intra-sujetos.

### *Tipos de contrastes*

**Desviación.** Compara la media de cada nivel (excepto una categoría de referencia) con la media de todos los niveles (media global). Los niveles del factor pueden colocarse en cualquier orden.

**Simple.** Compara la media de cada nivel con la media de un nivel especificado. Este tipo de contraste resulta útil cuando existe un grupo de control. Puede seleccionar la primera o la última categoría como referencia.

**Diferencia.** Compara la media de cada nivel (excepto el primero) con la media de los niveles anteriores (a veces también se denominan contrastes de Helmert inversos). (a veces también se denominan contrastes de Helmert inversos).

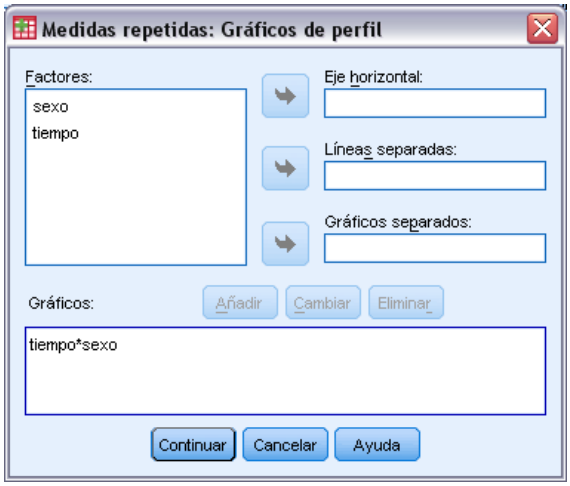
**Helmert.** Compara la media de cada nivel del factor (excepto el último) con la media de los niveles siguientes.

**Repetidas.** Compara la media de cada nivel (excepto el último) con la media del nivel siguiente.

**Polinómico.** Compara el efecto lineal, cuadrático, cúbico, etc. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, y así sucesivamente. Estos contrastes se utilizan a menudo para estimar las tendencias polinómicas.

## *MLG Medidas repetidas: Gráficos de perfil*

Figura 3-5
*Cuadro de diálogo Medidas repetidas: Gráficos de perfil*

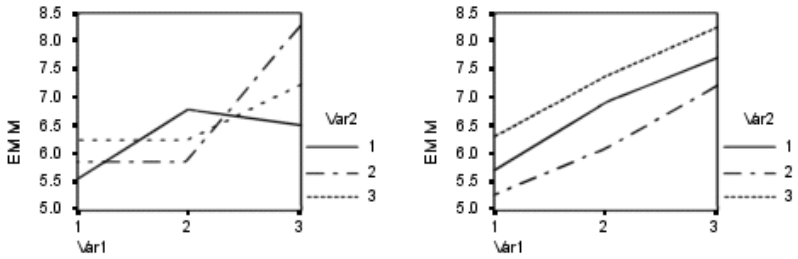



Los gráficos de perfil (gráficos de interacción) sirven para comparar las medias marginales en el modelo. Un gráfico de perfil es un gráfico de líneas en el que cada punto indica la media marginal estimada de una variable dependiente (corregida respecto a las covariables) en un nivel de un

factor. Los niveles de un segundo factor se pueden utilizar para generar líneas diferentes. Cada nivel en un tercer factor se puede utilizar para crear un gráfico diferente. Todos los factores están disponibles para los gráficos. Los gráficos de perfil se crean para cada variable dependiente. Es posible utilizar tanto los factores inter-sujetos como los intra-sujetos en los gráficos de perfil.

Un gráfico de perfil de un factor muestra si las medias marginales estimadas aumentan o disminuyen a través de los niveles. Para dos o más factores, las líneas paralelas indican que no existe interacción entre los factores, lo que significa que puede investigar los niveles de un único factor. Las líneas no paralelas indican una interacción.

Figura 3-6
*Gráfico no paralelo (izquierda) y gráfico paralelo (derecha)*




Después de especificar un gráfico mediante la selección de los factores del eje horizontal y, de manera opcional, los factores para distintas líneas y gráficos, el gráfico deberá añadirse a la lista de gráficos.

## MLG Medidas repetidas: Comparaciones post hoc

Figura 3-7
*Cuadro de diálogo Medidas repetidas: Comparaciones múltiples post hoc para las medias observadas*

**Pruebas de comparaciones múltiples post hoc** Una vez que se ha determinado que existen diferencias entre las medias, las pruebas de rango post hoc y las comparaciones múltiples por parejas permiten determinar qué medias difieren. Las comparaciones se realizan sobre valores sin corregir. Estas pruebas no están disponibles si no existen factores inter-sujetos y las pruebas de comparación múltiple post hoc se realizan para la media a través de los niveles de los factores intra-sujetos.

Las pruebas de diferencia honestamente significativa de Tukey y de Bonferroni son pruebas de comparación múltiple muy utilizadas. La **prueba de Bonferroni**, basada en el estadístico *t* de Student, corrige el nivel de significación observado por el hecho de que se realizan comparaciones múltiples. La **prueba t de Sidak** también corrige el nivel de significación y da lugar a límites más estrechos que los de Bonferroni. La **prueba de diferencia honestamente significativa de Tukey** utiliza el estadístico del rango estudentizado para realizar todas las comparaciones por pares entre los grupos y establece la tasa de error por experimento como la tasa de error para el conjunto de todas las comparaciones por pares. Cuando se contrasta un gran número de pares de medias, la prueba de la diferencia honestamente significativa de Tukey es más potente que la prueba de Bonferroni. Para un número reducido de pares, Bonferroni es más potente.

**GT2 de Hochberg** es similar a la prueba de la diferencia honestamente significativa de Tukey, pero se utiliza el módulo máximo estudentizado. La prueba de Tukey suele ser más potente. La **prueba de comparación por parejas de Gabriel** también utiliza el módulo máximo estudentizado y es generalmente más potente que la GT2 de Hochberg cuando los tamaños de las casillas son desiguales. La prueba de Gabriel se puede convertir en liberal cuando los tamaños de las casillas varían mucho.

La **prueba t de comparación múltiple por parejas de Dunnett** compara un conjunto de tratamientos con una media de control simple. La última categoría es la categoría de control por defecto. Si lo desea, puede seleccionar la primera categoría. Asimismo, puede elegir una prueba unilateral o bilateral. Para comprobar que la media de cualquier nivel del factor (excepto la categoría de control) no es igual a la de la categoría de control, utilice una prueba bilateral. Para contrastar si la media en cualquier nivel del factor es menor que la de la categoría de control, seleccione < Control. Asimismo, para contrastar si la media en cualquier nivel del factor es mayor que la de la categoría de control, seleccione > Control.

Ryan, Einot, Gabriel y Welsch (R-E-G-W) desarrollaron dos pruebas de rangos múltiples por pasos. Los procedimientos múltiples por pasos (por tamaño de las distancias) contrastan en primer lugar si todas las medias son iguales. Si no son iguales, se contrasta la igualdad en los subconjuntos de medias. **R-E-G-W F** se basa en una prueba *F* y **R-E-G-W Q** se basa en un rango estudentizado. Estas pruebas son más potentes que la prueba de rangos múltiples de Duncan y Student-Newman-Keuls (que también son procedimientos múltiples por pasos), pero no se recomiendan para tamaños de casillas desiguales.

Cuando las varianzas son desiguales, utilice **T2 de Tamhane** (prueba conservadora de comparación por parejas basada en una prueba *t*), **T3 de Dunnett** (prueba de comparación por parejas basada en el módulo máximo estudentizado), **prueba de comparación por parejasGames-Howell** (a veces liberal), o **C de Dunnett** (prueba de comparación por parejas basada en el rango estudentizado).

La **prueba de rango múltiple de Duncan**, Student-Newman-Keuls (**S-N-K**) y **b de Tukey** son pruebas de rango que asignan rangos a medias de grupo y calculan un valor de rango. Estas pruebas no se utilizan con la misma frecuencia que las pruebas anteriormente mencionadas.

La **prueba t de Waller-Duncan** utiliza la aproximación bayesiana. Esta prueba de rango emplea la media armónica del tamaño muestral cuando los tamaños muestrales no son iguales.

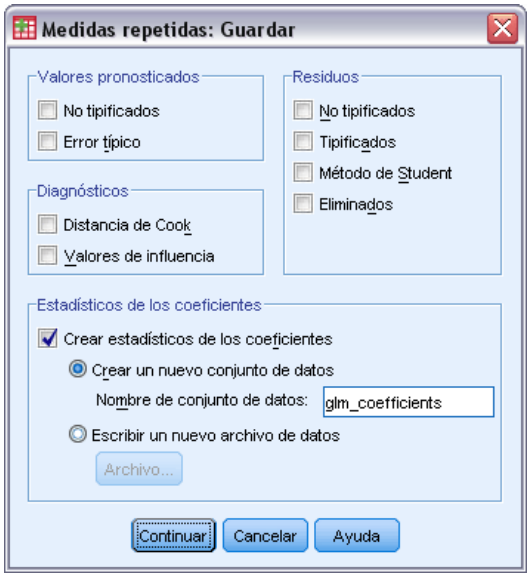
El nivel de significación de la prueba de **Scheffé** está diseñado para permitir todas las combinaciones lineales posibles de las medias de grupo que se van a contrastar, no sólo las comparaciones por parejas disponibles en esta función. El resultado es que la prueba de Scheffé es normalmente más conservadora que otras pruebas, lo que significa que se precisa una mayor diferencia entre las medias para la significación.

La prueba de comparación múltiple por parejas de la diferencia menos significativa (**DMS**) es equivalente a varias pruebas *t* individuales entre todos los pares de grupos. La desventaja de esta prueba es que no se realiza ningún intento de corregir el nivel crítico para realizar las comparaciones múltiples.

**Pruebas mostradas.** Se proporcionan comparaciones por parejas para DMS, Sidak, Bonferroni, Games-Howell, T2 y T3 de Tamhane, *C* de Dunnett y T3 de Dunnett. También se facilitan subconjuntos homogéneos para S-N-K, *b* de Tukey, Duncan, R-E-G-W *F*, R-E-G-W *Q* y Waller. La prueba de la diferencia honestamente significativa de Tukey, la GT2 de Hochberg, la prueba de Gabriel y la prueba de Scheffé son pruebas de comparaciones múltiples y pruebas de rango.

# MLG medidas repetidas: Guardar

Figura 3-8
*Cuadro de diálogo MLG Medidas repetidas: Guardar*




Es posible guardar los valores pronosticados por el modelo, los residuos y las medidas relacionadas como variables nuevas en el Editor de datos. Muchas de estas variables se pueden utilizar para examinar supuestos sobre los datos. Si desea almacenar los valores para utilizarlos en otra sesión de IBM® SPSS® Statistics, guárdelos en el archivo de datos actual.

**Valores pronosticados.** Son los valores que predice el modelo para cada caso.

- **No tipificados.** Valor pronosticado por el modelo para la variable dependiente.
- **Error típico.** Estimación de la desviación típica del valor promedio de la variable dependiente para los casos que tengan los mismos valores en las variables independientes.

**Diagnósticos.** Son medidas para identificar casos con combinaciones poco usuales de valores para los casos y las variables independientes que puedan tener un gran impacto en el modelo. Las opciones disponibles incluyen la distancia de Cook y los valores de influencia no centrados.

- **Distancia 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. Una Distancia de Cook grande indica que la exclusión de ese caso del cálculo de los estadísticos de regresión hará variar substancialmente los coeficientes.
- **Valores de influencia.** Valores de influencia no centrados. La influencia relativa de una observación en el ajuste del modelo.

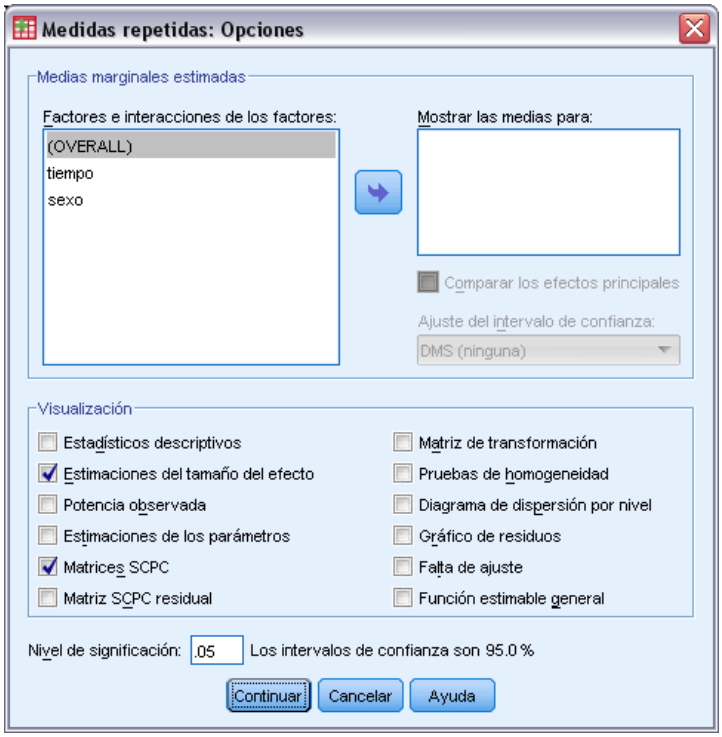
**Residuos.** Un residuo no tipificado es el valor real de la variable dependiente menos el valor pronosticado por el modelo. También se encuentran disponibles residuos eliminados, estudentizados y tipificados.

- **No tipificados.** Diferencia entre un valor observado y el valor pronosticado por el modelo.
- **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.
- **Método de Student.** Residuo dividido por una estimación de su desviación típica que varía de caso en caso, dependiendo de la distancia de los valores de cada caso en las variables independientes respecto a las medias en las variables independientes.
- **Eliminados.** Residuo para un caso cuando éste se excluye del cálculo de los coeficientes de la regresión. Es igual a la diferencia entre el valor de la variable dependiente y el valor pronosticado corregido.

**Estadísticos de los coeficientes.** Guarda una matriz varianza-covarianza o una matriz de las estimaciones de los parámetros en un conjunto de datos o archivo de datos. Asimismo, para cada variable dependiente habrá una fila de estimaciones de los parámetros, una fila de valores de significación para los estadísticos *t* correspondientes a las estimaciones de los parámetros y una fila de grados de libertad de los residuos. En un modelo multivariante, existen filas similares para cada variable dependiente. Si lo desea, puede usar estos datos matriciales en otros procedimientos que lean archivos matriciales. Los conjuntos de datos están disponibles para su uso posterior durante la misma sesión, pero no se guardarán como archivos a menos que se hayan guardado explícitamente antes de que finalice la sesión. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables.

# MLG Medidas repetidas: Opciones

Figura 3-9
*Cuadro de diálogo MLG Medidas repetidas: Opciones*




Este cuadro de diálogo contiene estadísticos opcionales. Los estadísticos se calculan utilizando un modelo de efectos fijos.

**Medias marginales estimadas.** Seleccione los factores e interacciones para los que desee obtener estimaciones de las medias marginales de la población en las casillas. Estas medias se corrigen respecto a las covariables, si las hay. Se pueden seleccionar tanto factores intra-sujetos como inter-sujetos.

- **Comparar los efectos principales.** Proporciona comparaciones por parejas no corregidas entre las medias marginales estimadas para cualquier efecto principal del modelo, tanto para los factores inter-sujetos como para los intra-sujetos. Este elemento sólo se encuentra disponible si los efectos principales están seleccionados en la lista Mostrar las medias para.
- **Ajuste del intervalo de confianza.** Seleccione un ajuste de diferencia menor significativa (DMS), Bonferroni o Sidak para los intervalos de confianza y la significación. Este elemento sólo estará disponible si se selecciona Comparar los efectos principales.

**Mostrar.** Seleccione Estadísticos descriptivos para generar medias observadas, desviaciones típicas y frecuencias para cada variable dependiente en todas las casillas. La opción Estimaciones del tamaño del efecto ofrece un valor parcial de eta-cuadrado para cada efecto y cada estimación de parámetros. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor. Seleccione Potencia observada para obtener la potencia de la prueba cuando la hipótesis alternativa se ha establecido basándose en el valor observado. Seleccione Estimaciones de los parámetros para generar las estimaciones de los parámetros, los errores típicos, las pruebas *t*, los

intervalos de confianza y la potencia observada para cada prueba. Se pueden mostrar Matrices SCPC de error y de hipótesis y la Matriz SCPC residual más la prueba de esfericidad de Bartlett de la matriz de covarianzas residual.

Las pruebas de homogeneidad producen la prueba de homogeneidad de varianzas de Levene para cada variable dependiente en todas las combinaciones de nivel de los factores inter-sujetos sólo para factores inter-sujetos. Asimismo, las pruebas de homogeneidad incluyen la prueba *M* de Box sobre la homogeneidad de las matrices de covarianza de las variables dependientes a lo largo de todas las combinaciones de niveles de los factores inter-sujetos. Las opciones de diagramas de dispersión por nivel y gráfico de los residuos son útiles para comprobar los supuestos sobre los datos. Estos elementos no estarán activado si no hay factores. Seleccione Gráficos de los residuos para generar un gráfico de los residuos observados respecto a los pronosticados respecto a los tipificados para cada variable dependiente. Estos gráficos son útiles para investigar el supuesto de varianzas iguales. Seleccione la Prueba de falta de ajuste para comprobar si el modelo puede describir de forma adecuada la relación entre la variable dependiente y las variables independientes. La función estimable general permite construir pruebas de hipótesis personales basadas en la función estimable general. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función estimable general.

**Nivel de significación.** Puede que le interese corregir el nivel de significación usado en las pruebas post hoc y el nivel de confianza empleado para construir intervalos de confianza. El valor especificado también se utiliza para calcular la potencia observada para la prueba. Si especifica un nivel de significación, el cuadro de diálogo mostrará el nivel asociado de los intervalos de confianza.

## *Funciones adicionales del comando GLM*

Estas funciones se pueden aplicar a los análisis univariados, multivariados o de medidas repetidas. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar efectos anidados en el diseño (utilizando el subcomando `DESIGN`).
- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando `TEST`).
- Especificar contrastes múltiples (utilizando el subcomando `CONTRAST`).
- Incluir los valores perdidos definidos por el usuario (utilizando el subcomando `MISSING`).
- Especificar criterios EPS (mediante el subcomando `CRITERIA`).
- Construir una matriz**L**, una matriz **M** o una matriz **K** (utilizando los subcomandos `LMATRIX`, `MMATRIX` y `KMATRIX`).
- Especificar una categoría de referencia intermedia (utilizando el subcomando `CONTRAST` para los contrastes de desviación o simples).
- Especificar la métrica para los contrastes polinómicos (utilizando el subcomando `CONTRAST`).
- Especificar términos de error para las comparaciones post hoc (utilizando el subcomando `POSTHOC`).
- Calcular medias marginales estimadas para cualquier factor o interacción entre los factores en la lista de factores (utilizando el subcomando `EMMEANS`).
- Especificar nombres para las variables temporales (utilizando el subcomando `SAVE`).

- Construir un archivo de datos matricial de correlaciones (utilizando el subcomando `OUTFILE`).
- Construir un archivo de datos matricial que contenga estadísticos de la tabla de ANOVA inter-sujetos (utilizando el subcomando `OUTFILE`).
- Guardar la matriz del diseño en un nuevo archivo de datos (utilizando el subcomando `OUTFILE`).

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

Capítulo
4

# Análisis de componentes de la varianza

El procedimiento Componentes de la varianza, para modelos de efectos mixtos, estima la contribución de cada efecto aleatorio a la varianza de la variable dependiente. Este procedimiento resulta de particular interés para el análisis de modelos mixtos, como los diseños split-plot, los diseños de medidas repetidas univariados y los diseños de bloques aleatorios. Al calcular las componentes de la varianza, se puede determinar dónde centrar la atención para reducir la varianza.

Se dispone de cuatro métodos diferentes para estimar las componentes de la varianza: estimador mínimo no cuadrático insesgado (EMNCI, MINQUE), análisis de varianza (ANOVA), máxima verosimilitud (MV, ML) y máxima verosimilitud restringida (MVR, RML). Se dispone de diversas especificaciones para los diferentes métodos.

Los resultados por defecto para todos los métodos incluyen las estimaciones de componentes de la varianza. Si se usa el método MV o el método MVR, se mostrará también una tabla con la matriz de covarianzas asintótica. Otros resultados disponibles incluyen una tabla de ANOVA y las medias cuadráticas esperadas para el método ANOVA, y la historia de iteraciones para los métodos MV y MVR. El procedimiento Componentes de la varianza es totalmente compatible con el procedimiento MLG Factorial general.

La opción Ponderación MCP permite especificar una variable usada para aplicar a las observaciones diferentes ponderaciones para un análisis ponderado; por ejemplo, para compensar las variaciones de precisión de las medidas.

**Ejemplo.** En una escuela agrícola, se mide el aumento de peso de los cerdos de seis camadas diferentes después de un mes. La variable camada es un factor aleatorio con seis niveles. Las seis camadas estudiadas son una muestra aleatoria de una amplia población de camadas de cerdos. El investigador deduce que la varianza del aumento de peso se puede atribuir a la diferencia entre las camadas más que a la diferencia entre los cerdos de una misma camada.

**Datos.** La variable dependiente es cuantitativa. Los factores son categóricos; pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Pueden tener valores numéricos o valores de cadena de hasta ocho bytes. Al menos uno de los factores debe ser aleatorio. Es decir, los niveles del factor deben ser una muestra aleatoria de los posibles niveles. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente.

**Supuestos.** Todos los métodos suponen que los parámetros del modelo para un efecto aleatorio tienen de media cero y varianzas constantes finitas y no están correlacionados mutuamente. Los parámetros del modelo para diferentes efectos aleatorios son también independientes.

El término residual también tiene una media de cero y una varianza constante finita. No tiene correlación con respecto a los parámetros del modelo de cualquier efecto aleatorio. Se asume que los términos residuales de diferentes observaciones no están correlacionados.

Basándose en estos supuestos, las observaciones del mismo nivel de un factor aleatorio están correlacionadas. Este hecho distingue un modelo de componentes de la varianza a partir de un modelo lineal general.

ANOVA y EMNCI no requieren supuestos de normalidad. Ambos son robustos a las desviaciones moderadas del supuesto de normalidad.

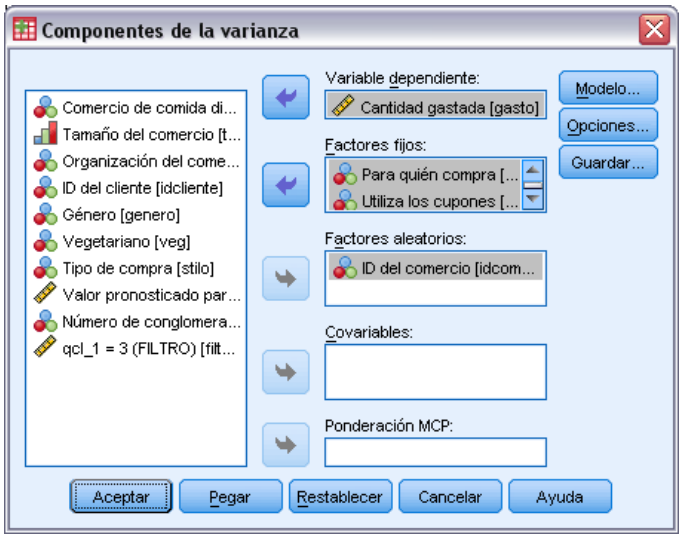
MV y MVR requieren que el parámetro del modelo y el término residual se distribuyan de forma normal.

**Procedimientos relacionados.** Use el procedimiento Explorar para examinar los datos antes de realizar el análisis de componentes de la varianza. Para contrastar hipótesis, utilice MLG Factorial general, MLG Multivariado y MLG Medidas repetidas.

### *Para obtener un análisis de las componentes de la varianza*

- En los menús, seleccione:
  Analizar > Modelo lineal general > Componentes de la varianza...

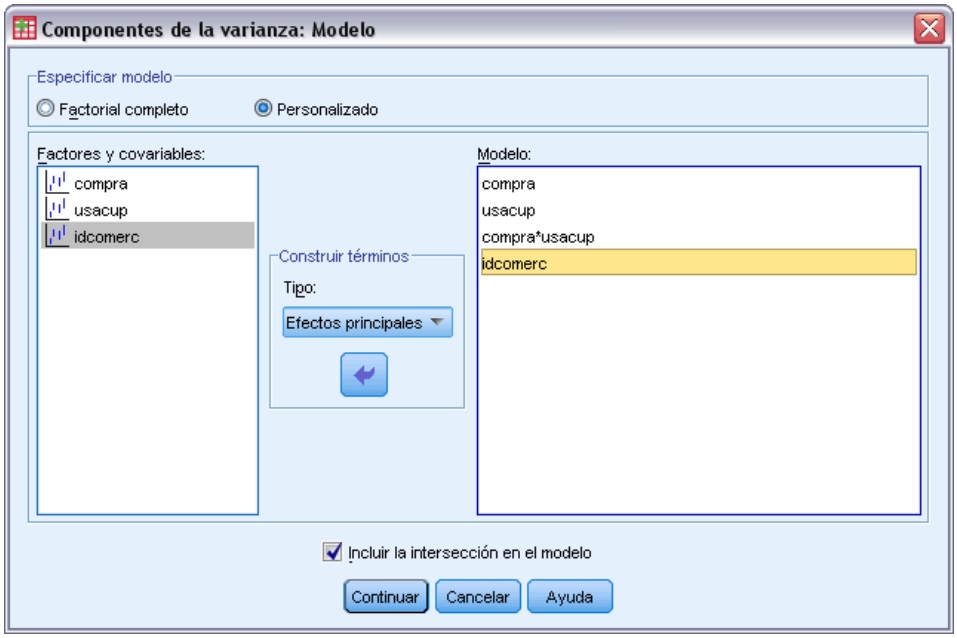
Figura 4-1
*Cuadro de diálogo Componentes de la varianza*




- Seleccione una variable dependiente.
- Seleccione variables para Factores fijos, Factores aleatorios y Covariables, en función de los datos. Para especificar una variable de ponderación, utilice Ponderación MCP.

# Componentes de la varianza: Modelo

Figura 4-2
*Cuadro de diálogo Componentes de la varianza: Modelo*




**Especificar modelo.** Un modelo factorial completo contiene todos los efectos principales del factor, todos los efectos principales de las covariables y todas las interacciones factor por factor. No contiene interacciones de covariable. Seleccione Personalizado para especificar sólo un subconjunto de interacciones o para especificar interacciones factor por covariable. Indique todos los términos que desee incluir en el modelo.

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

**Modelo.** El modelo depende de la naturaleza de los datos. Después de seleccionar Personalizado, puede elegir los efectos principales y las interacciones que sean de interés para el análisis. El modelo debe contener un factor aleatorio.

**Incluir la intersección en el modelo.** Normalmente se incluye la intersección en el modelo. Si supone que los datos pasan por el origen, puede excluir la intersección.

## Construir términos

Para las covariables y los factores seleccionados:

**Interacción.** Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el método por defecto.

**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.

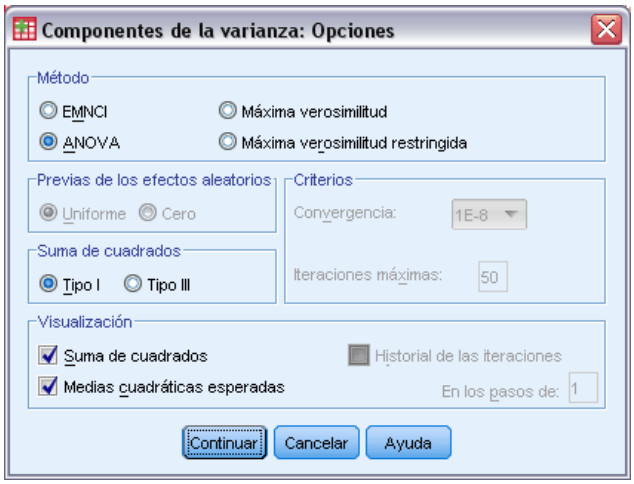
**Todas de 3.** Crea todas las interacciones triples posibles de las variables seleccionadas.

**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.

## Componentes de la varianza: Opciones

Figura 4-3
*Cuadro de diálogo Componentes de la varianza: Opciones*




**Método.** Puede seleccionar uno de los cuatro métodos para estimar las componentes de la varianza.

- EMNCI (estimador mínimo no cuadrático insesgado) produce estimaciones que son invariables con respecto a los efectos fijos. Si los datos se distribuyen normalmente y las estimaciones son correctas, este método produce la varianza inferior entre todos los estimadores insesgados. Puede seleccionar un método para las ponderaciones previas de los efectos aleatorios.
- ANOVA (análisis de varianza) calcula las estimaciones insesgadas utilizando las sumas de cuadrados de Tipo I o Tipo III para cada efecto. El método ANOVA a veces produce estimaciones de varianza negativas, que pueden indicar un modelo erróneo, un método de estimación inadecuado o la necesidad de más datos.
- Máxima verosimilitud (MV) genera estimaciones que serán lo más coherente posible con los datos observados realmente, utilizando iteraciones. Estas estimaciones pueden estar sesgadas. Este método es asintóticamente normal. Las estimaciones MV y MVR son invariables a la traslación. Este método no tiene en cuenta los grados de libertad utilizados para estimar los efectos fijos.
- Las estimaciones de máxima verosimilitud restringida (MVR) reducen las estimaciones ANOVA para muchos (si no todos) los casos de datos equilibrados. Puesto que este método se corrige respecto a los efectos fijos, deberá dar errores típicos menores que el método MV. Este método tiene en consideración los grados de libertad utilizados para estimar los efectos fijos.

**Previas de los efectos aleatorios.** Uniforme implica que todos los efectos aleatorios y el término residual tienen un impacto igual en las observaciones. El esquema Cero equivale a asumir varianzas de efecto aleatorio cero. Sólo se encuentra disponible para el método EMNCI.

**Suma de cuadrados** Las sumas de cuadrados de Tipo I se utilizan para el modelo jerárquico, el cual es empleado con frecuencia en las obras sobre componentes de la varianza. Si selecciona Tipo III, que es el valor por defecto en MLG, las estimaciones de la varianza podrán utilizarse en MLG Factorial general para contrastar hipótesis con sumas de cuadrados de Tipo III. Sólo se encuentra disponible para el método ANOVA.

**Criterios.** Puede especificar el criterio de convergencia y el número máximo de iteraciones. Sólo se encuentra disponible para los métodos MV o MVR.

**Mostrar.** Para el método ANOVA, puede seleccionar mostrar sumas de cuadrados y medias cuadráticas esperadas. Si selecciona el método de Máxima verosimilitud o el de Máxima verosimilitud restringida, puede mostrar una historia de las iteraciones.

## *Sumas de cuadrados (Componentes de la varianza)*

Para el modelo, puede elegir un tipo de suma de cuadrados. El Tipo III es el más utilizado y es el tipo por defecto.

**Tipo I.** Este método también se conoce como el método de descomposición jerárquica de la suma de cuadrados. Cada término se corrige sólo respecto al término que le precede en el modelo. El método Tipo I para la obtención de sumas de cuadrados se utiliza normalmente para:

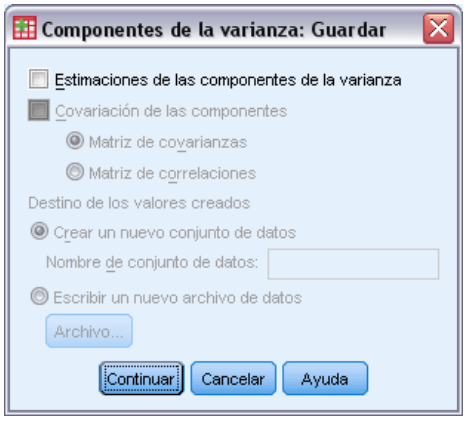
- Un modelo ANOVA equilibrado en el que se especifica cualquier efecto principal antes de cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se especifica antes de cualquier efecto de interacción de segundo orden, y así sucesivamente.
- Un modelo de regresión polinómica en el que se especifica cualquier término de orden inferior antes que cualquier término de orden superior.
- Un modelo puramente anidado en el que el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente. Esta forma de anidamiento solamente puede especificarse utilizando la sintaxis.

**Tipo III.** Es el método por defecto. Este método calcula las sumas de cuadrados de un efecto del diseño como las sumas de cuadrados corregidas respecto a cualquier otro efecto que no lo contenga y ortogonales a cualquier efecto (si existe) que lo contenga. Las sumas de cuadrados de Tipo III tienen una gran ventaja por ser invariables respecto a las frecuencias de casilla, siempre que la forma general de estimabilidad permanezca constante. Así, este tipo de sumas de cuadrados se considera a menudo útil para un modelo no equilibrado sin casillas perdidas. En un diseño factorial sin casillas perdidas, este método equivale a la técnica de cuadrados ponderados de las medias de Yates. El método Tipo III para la obtención de sumas de cuadrados se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en Tipo I.
- Cualquier modelo equilibrado o desequilibrado sin casillas vacías.

## Componentes de la varianza: Guardar en archivo nuevo

Figura 4-4
*Cuadro de diálogo Componentes de la varianza: Guardar en archivo nuevo*




Se pueden guardar algunos resultados de este procedimiento en un nuevo archivo de datos IBM® SPSS® Statistics.

**Estimaciones de las componentes de la varianza.** Guarda las estimaciones de las componentes de la varianza y las etiquetas de estimación en un archivo de datos o conjunto de datos. Se puede utilizar para calcular más estadísticos o en otros análisis de los procedimientos MLG. Por ejemplo, se pueden usar para calcular intervalos de confianza o para contrastar hipótesis.

**Covariación de las componentes.** Guarda una matriz varianza-covarianza o una matriz de correlaciones en un archivo de datos o conjunto de datos. Sólo está disponible si se han especificado los métodos de máxima verosimilitud o máxima verosimilitud restringida.

**Destino de los valores creados.** Permite especificar un nombre para un conjunto de datos o para un archivo externo que contenga las estimaciones de las componentes de la varianza y/o la matriz. Los conjuntos de datos están disponibles para su uso posterior durante la misma sesión, pero no se guardarán como archivos a menos que se hayan guardado explícitamente antes de que finalice la sesión. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables.

Se puede utilizar el comando `MATRIX` para extraer los datos que necesite del archivo de datos y después calcular los intervalos de confianza o realizar pruebas.

## Funciones adicionales del comando VARCOMP

Con el lenguaje de sintaxis de comandos también podrá:

- Especificar efectos anidados en el diseño (utilizando el subcomando `DESIGN`).
- Incluir los valores perdidos definidos por el usuario (utilizando el subcomando `MISSING`).
- Especificar criterios EPS (mediante el subcomando `CRITERIA`).

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

Capítulo
5

# Modelos lineales mixtos

El procedimiento Modelos lineales mixtos amplía el modelo lineal general de manera que los datos puedan presentar variabilidad corelacionada y no constante. El modelo lineal mixto proporciona, por tanto, la flexibilidad necesaria para modelar no sólo las medias sino también las varianzas y covarianzas de los datos.

El procedimiento Modelos lineales mixtos es asimismo una herramienta flexible para ajustar otros modelos que puedan ser formulados como modelos lineales mixtos. Dichos modelos incluyen los modelos multinivel, los modelos lineales jerárquicos y los modelos con coeficientes aleatorios.

**Ejemplo.** Una cadena de tiendas de comestibles está interesada en los efectos de varios vales en el gasto de los clientes. Se toma una muestra aleatoria de los clientes habituales para observar el gasto de cada cliente durante 10 semanas. Cada semana se envía por correo un vale distinto a los clientes. Los modelos lineales mixtos se utilizan para estimar el efecto de los distintos vales en el gasto, a la vez que se corrige respecto a la correlación debida a las observaciones repetidas de cada sujeto durante las 10 semanas.

**Métodos.** Estimación de máxima verosimilitud (MV) y máxima verosimilitud restringida (MVR).

**Estadísticos.** Estadísticos descriptivos: tamaños de las muestras, medias y desviaciones típicas de la variable dependiente y las covariables para cada combinación de niveles de los factores. Información de los niveles del factor: valores ordenados de los niveles de cada factor y las frecuencias correspondientes. Asimismo, las estimaciones de los parámetros y los intervalos de confianza para los efectos fijos y las pruebas de Wald y los intervalos de confianza para los parámetros de las matrices de covarianzas. Pueden emplearse las sumas de cuadrados de Tipo I y Tipo III para evaluar diferentes hipótesis. Tipo III es el valor por defecto.

**Datos.** La variable dependiente debe ser cuantitativa. Los factores deben ser categóricos y pueden tener valores numéricos o valores de cadena. Las covariables y la variable de ponderación deben ser cuantitativas. Las variables de sujetos y repetidas pueden ser de cualquier tipo.

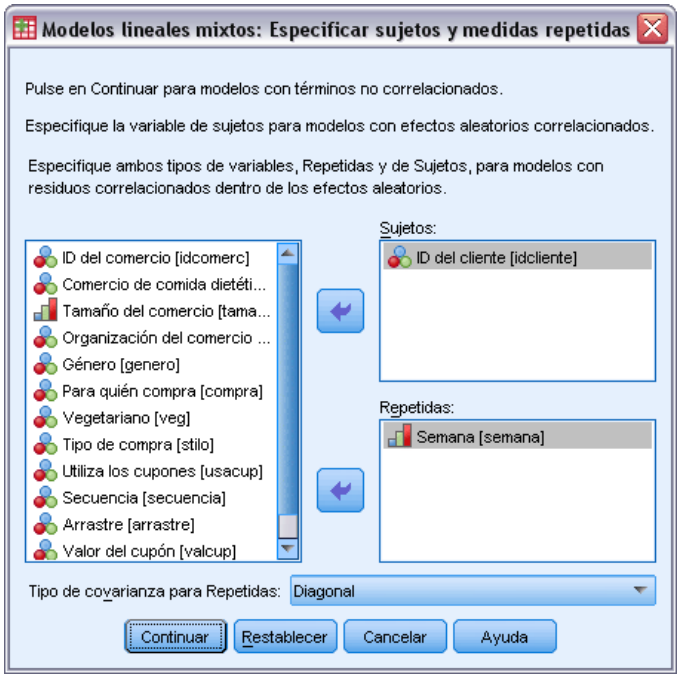
**Supuestos.** Se supone que la variable dependiente está relacionada linealmente con los factores fijos, los factores aleatorios y las covariables. Los efectos fijos modelan la media de la variable dependiente. Los efectos aleatorios modelan la estructura de las covarianzas de la variable dependiente. Los efectos aleatorios múltiples se consideran independientes entre sí y se calculan por separado las matrices de covarianzas de cada uno de ellos; sin embargo, se puede establecer una correlación entre los términos del modelo especificados para el mismo efecto aleatorio. Las medidas repetidas modelan la estructura de las covarianzas de los residuos. Se asume además que la variable dependiente procede de una distribución normal.

**Procedimientos relacionados.** Use el procedimiento Explorar para examinar los datos antes de realizar un análisis. Si no cree que haya una variabilidad correlacionada o no constante, puede usar alternativamente el procedimiento MLG Univariante o MLG Medidas repetidas. Alternativamente, puede usar el procedimiento Análisis de componentes de la varianza en caso de que los efectos aleatorios tengan una estructura de covarianzas en los componentes de la varianza y no haya medidas repetidas.

### *Obtención de un análisis de Modelos lineales mixtos*

- En los menús, seleccione:
  Analizar > Modelos mixtos > Lineal...

Figura 5-1
*Cuadro de diálogo Modelos lineales mixtos: Especificar sujetos y medidas repetidas*




- Si lo desea, puede seleccionar una o más variables de sujetos.
- Si lo desea, puede seleccionar una o más variables repetidas.
- Si lo desea, puede seleccionar una estructura de covarianza residual.
- Pulse en Continuar.

Figura 5-2
*Cuadro de diálogo Modelos lineales mixtos*


- Seleccione una variable dependiente.
- Seleccione al menos un factor o covariable.
- Pulse en Fijos o Aleatorios y especifique al menos un modelo de efectos fijos o aleatorios.

  Si lo desea, seleccione una variable de ponderación.

## *Modelos lineales mixtos: Selección de las variables de Sujetos/Repetidas*

Este cuadro de diálogo le permite seleccionar variables que definen sujetos y observaciones repetidas, y elegir una estructura de covarianzas para los residuos. Consulte Figura 5-1 el p. 38.

**Sujetos.** Un sujeto es una unidad de observación, la cual se puede considerar independiente de otros sujetos. Por ejemplo, en un estudio médico, las lecturas de la presión sanguínea de un paciente se pueden considerar independientes de las lecturas de otros pacientes. La definición de los sujetos es particularmente importante cuando se dan medidas repetidas para cada sujeto y desea modelar la correlación entre estas observaciones. Por ejemplo, cabe esperar que estén correlacionadas las lecturas de la presión sanguínea de un único paciente en una serie de visitas consecutivas al médico.

Los sujetos se pueden definir además mediante la combinación de los niveles de los factores de múltiples variables; por ejemplo, puede especificar el *Sexo* y la *Categoría de edad* como variables de sujetos para modelar la creencia de que los *hombres de más de 65 años* son similares entre sí, pero independientes de los *hombres de menos de 65 años* y de las *mujeres*.

Todas las variables especificadas en la lista Sujetos se usan con el fin de definir los sujetos para la estructura de la covarianza residual. Puede usar todas o algunas de las variables que definen los sujetos para la estructura de la covarianza de los efectos aleatorios.

**Repetidas.** Las variables especificadas en esta lista se usan para identificar las observaciones repetidas. Por ejemplo, una única variable *Semana* puede identificar las 10 semanas de observaciones de un estudio médico o se pueden usar *Mes* y *Día* para identificar las observaciones diarias realizadas a lo largo de un año.

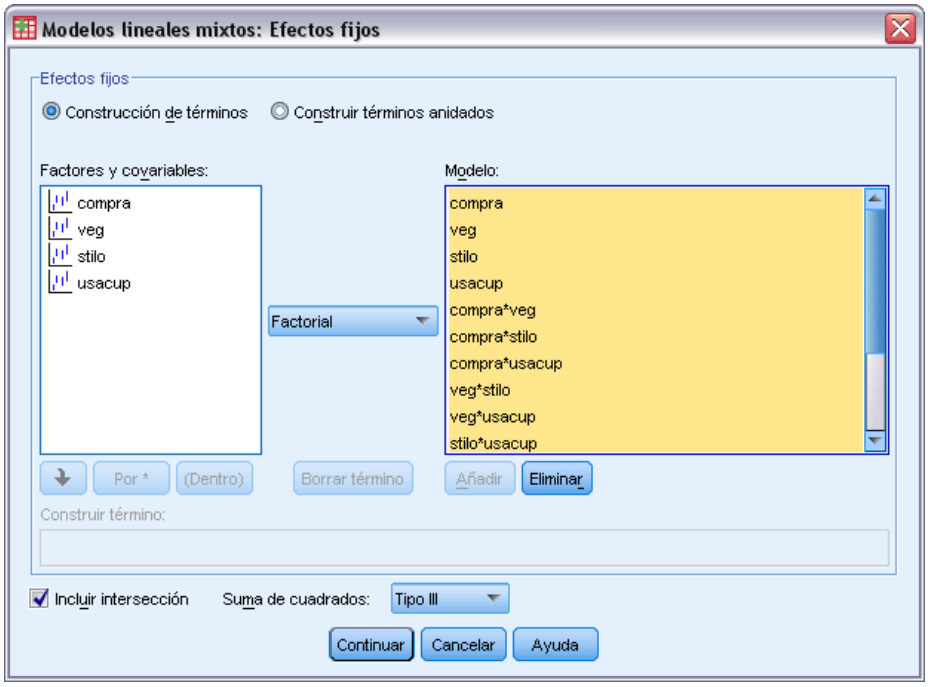
**Tipo de covarianza para Repetidas.** Especifica la estructura de la covarianza para los residuos. Las estructuras disponibles son las siguientes:

- Dependencia Ante: Primer orden
- AR(1).
- AR(1): Heterogénea
- ARMA(1,1).
- Simetría compuesta
- Simetría compuesta: Métrica de correlación
- Simetría compuesta: Heterogénea
- Diagonal
- Factor analítico: Primer orden
- Factor analítico: Primer orden, Heterogéneo
- Huynh-Feldt
- Identidad escalada
- Toeplitz
- Toeplitz: Heterogénea
- Sin estructura
- Sin estructura: Correlaciones

Si desea obtener más información, consulte el tema Estructuras de covarianza en el apéndice B el p. 166.

# Efectos fijos de los Modelos lineales mixtos

Figura 5-3
*Cuadro de diálogo Efectos fijos de los modelos lineales mixtos*




**Efectos fijos.** No existe un modelo por defecto, por lo que debe especificar de forma explícita los efectos fijos. Puede elegir entre términos anidados o no anidados.

**Incluir intersección.** La intersección se incluye normalmente en el modelo. Si asume que los datos pasan por el origen, puede excluir la intersección.

**Suma de cuadrados** Determina el método para calcular las sumas de cuadrados. En el caso de los modelos sin casillas perdidas, el método Tipo III es por lo general el más utilizado.

## Construir términos no anidados

Para las covariables y los factores seleccionados:

**Factorial.** Crea todas las interacciones y efectos principales posibles para las variables seleccionadas. Este es el método por defecto.

**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.

## *Construir términos anidados*

En este procedimiento, puede construir términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento del gasto de sus clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción o añadir varios niveles de anidación al término anidado.

**Limitaciones.** Existen las siguientes restricciones para los términos anidados:

- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A*A*.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A(A)*.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si *A* es un factor y *X* es una covariable, no es válido especificar *A(X)*.

## *Suma de cuadrados*

Para el modelo, puede elegir un tipo de suma de cuadrados. El Tipo III es el más utilizado y es el tipo por defecto.

**Tipo I.** Este método también se conoce como el método de descomposición jerárquica de la suma de cuadrados. Cada término se corrige sólo para el término que le precede en el modelo. El método Tipo I para la obtención de sumas de cuadrados se utiliza normalmente para:

- Un modelo ANOVA equilibrado en el que se especifica cualquier efecto principal antes de cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se especifica antes de cualquier efecto de interacción de segundo orden, y así sucesivamente.
- Un modelo de regresión polinómica en el que se especifica cualquier término de orden inferior antes que cualquier término de orden superior.
- Un modelo puramente anidado en el que el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente. Esta forma de anidamiento solamente puede especificarse utilizando la sintaxis.

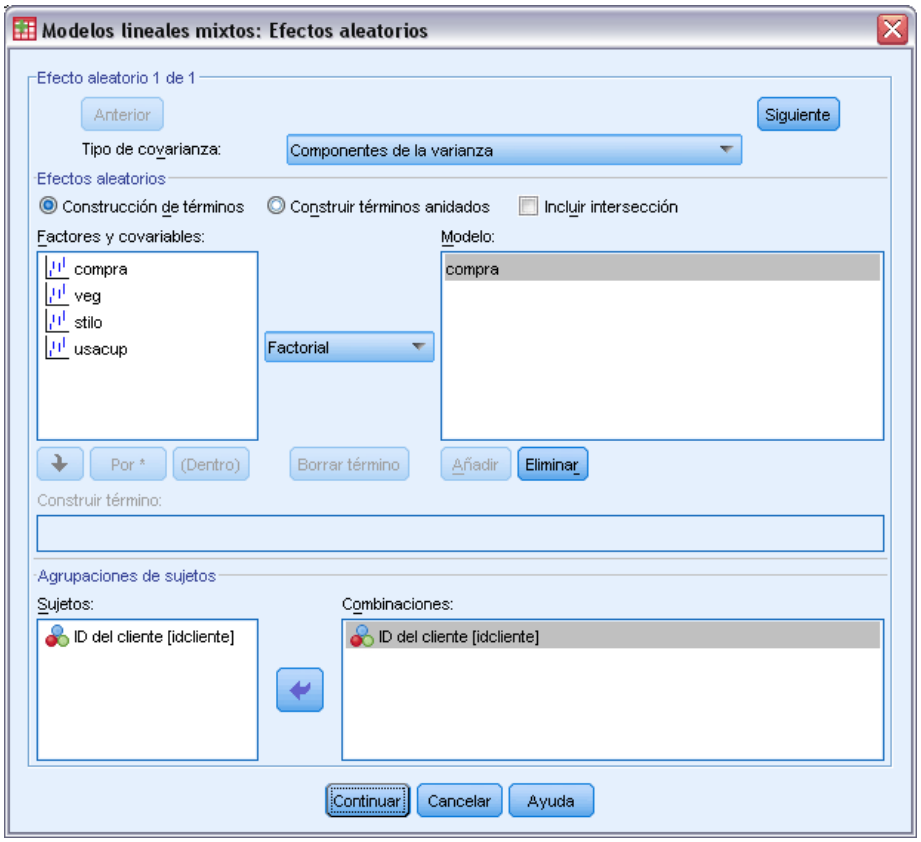
**Tipo III.** Es el método por defecto. Este método calcula las sumas de cuadrados de un efecto del diseño como las sumas de cuadrados corregidas respecto a cualquier otro efecto que no lo contenga y ortogonales a cualquier efecto (si existe) que lo contenga. Las sumas de cuadrados de Tipo III tienen una gran ventaja por ser invariables respecto a las frecuencias de casilla, siempre que la forma general de estimabilidad permanezca constante. Así, este tipo de sumas de cuadrados se suele considerar de gran utilidad para un modelo no equilibrado sin casillas perdidas. En un diseño factorial sin casillas perdidas, este método equivale a la técnica de cuadrados ponderados

de las medias de Yates. El método Tipo III para la obtención de sumas de cuadrados se utiliza normalmente para:

- Cualquiera de los modelos que aparecen en Tipo I.
- Cualquier modelo equilibrado o desequilibrado sin casillas vacías.

## Efectos aleatorios de los Modelos lineales mixtos

Figura 5-4
*Cuadro de diálogo Efectos aleatorios de los modelos lineales mixtos*




**Tipo de covarianza.** Le permite especificar la estructura de las covarianzas para el modelo de efectos aleatorios. Para cada efecto aleatorio se estima una matriz de covarianzas por separado. Las estructuras disponibles son las siguientes:

- Dependencia Ante: Primer orden
- AR(1).
- AR(1): Heterogénea
- ARMA(1,1).
- Simetría compuesta
- Simetría compuesta: Métrica de correlación
- Simetría compuesta: Heterogénea

- Diagonal
- Factor analítico: Primer orden
- Factor analítico: Primer orden, Heterogéneo
- Huynh-Feldt
- Identidad escalada
- Toeplitz
- Toeplitz: Heterogénea
- Sin estructura
- Sin estructura: Métrica de correlación
- Variance Components

Si desea obtener más información, consulte el tema Estructuras de covarianza en el apéndice B el p. 166.

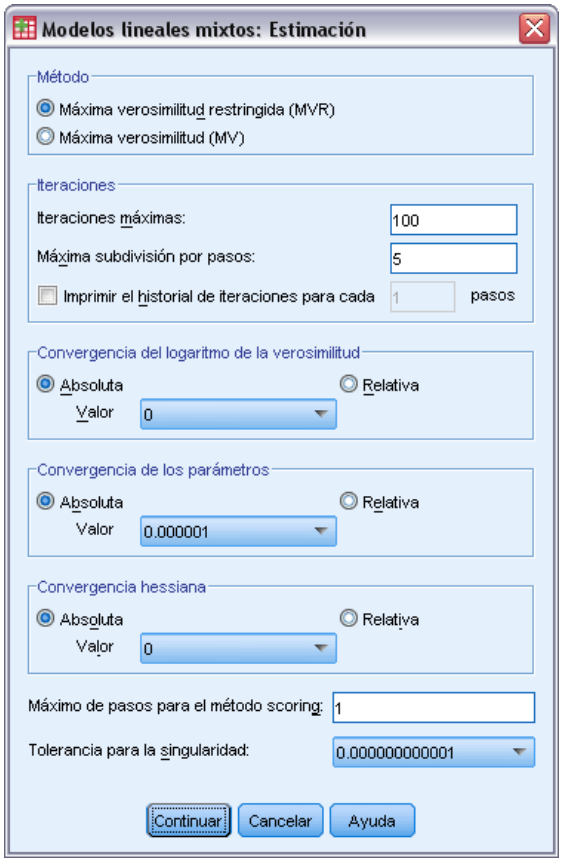
**Efectos aleatorios.** No existe un modelo por defecto, por lo que debe especificar de forma explícita los efectos aleatorios. Puede elegir entre términos anidados o no anidados. Puede asimismo incluir un término de intersección en el modelo de efectos aleatorios.

Puede especificar varios modelos de efectos aleatorios. Una vez construido el primer modelo, pulse en Siguiente para construir el siguiente modelo. Pulse en Anterior para desplazarse hacia atrás por los modelos existentes. Cada modelo de efecto aleatorio se supone que es independiente del resto de los modelos de efectos aleatorios; es decir, se calcularán diferentes matrices de covarianzas para cada uno de ellos. Se puede establecer una correlación entre los términos especificados en el mismo modelo de efectos aleatorios.

**Agrupaciones de sujetos.** Las variables incluidas son las seleccionadas como variables de sujetos en el cuadro de diálogo Selección de variables de sujetos/repetidas. Elija todas o algunas de las variables para definir los sujetos en el modelo de efectos aleatorios.

# Estimación de los Modelos lineales mixtos

Figura 5-5
*Cuadro de diálogo Estimación de modelos lineales mixtos*




**Método.** Seleccione la estimación de máxima verosimilitud o de máxima verosimilitud restringida.

**Iteraciones:**

- **Iteraciones máximas.** Especifique un número entero no negativo.
- **Máxima subdivisión por pasos.** En cada iteración, se reduce el tamaño del paso mediante un factor de 0,5 hasta que aumenta el logaritmo de la verosimilitud o se alcanza la máxima subdivisión por pasos. Especifique un número entero positivo.
- **Imprimir el historial de iteraciones para cada n pasos.** Muestra una tabla que incluye el valor de la función del logaritmo de la verosimilitud y las estimaciones de los parámetros cada *n* iteraciones, comenzando por la iteración 0 (las estimaciones iniciales). Si decide imprimir el historial de iteraciones, la última iteración se imprimirá siempre independientemente del valor de *n*.

**Convergencia del logaritmo de la verosimilitud.** Se asume la convergencia si el cambio absoluto o relativo en la función del logaritmo de la verosimilitud es inferior al valor especificado, el cual debe ser no negativo. Si el valor especificado es igual a 0, no se utiliza el criterio.

**Convergencia de los parámetros.** Se asume la convergencia si el cambio absoluto o relativo máximo en las estimaciones de los parámetros es inferior al valor especificado, el cual debe ser no negativo. Si el valor especificado es igual a 0, no se utiliza el criterio.

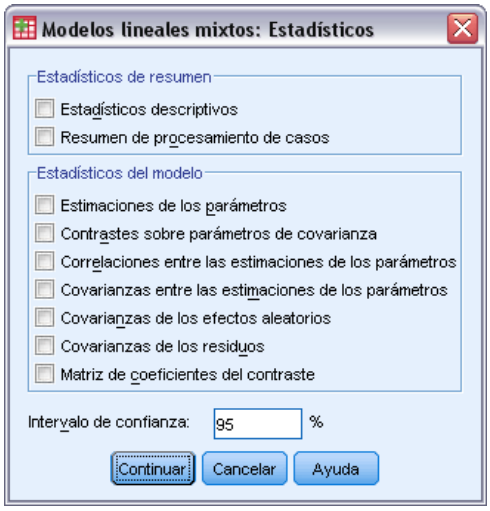
**Convergencia hessiana.** En el caso de la especificación Absoluta, se asume la convergencia si un estadístico basado en la hessiana es inferior al valor especificado. En el caso de la especificación Relativa, se asume la convergencia si el estadístico es inferior al producto del valor especificado y el valor absoluto del logaritmo de la verosimilitud. Si el valor especificado es igual a 0, no se utiliza el criterio.

**Máximo de pasos para el método scoring.** Solicita utilizar el algoritmo de puntuación de Fisher hasta el número de iteraciones *n*. Especifique un número entero positivo.

**Tolerancia para la singularidad.** Este valor se utiliza como tolerancia en la comprobación de la singularidad. Especifique un valor positivo.

## Estadísticos de Modelos lineales mixtos

Figura 5-6
*Cuadro de diálogo Estadísticos de modelos lineales mixtos*




**Estadísticos de resumen.** Genera tablas correspondientes a:

- **Estadísticos descriptivos.** Muestra los tamaños de las muestras, medias y desviaciones típicas de la variable dependiente y las covariables (si se especifican). Estos estadísticos se muestran para cada combinación de niveles de los factores.
- **Resumen de procesamiento de casos.** Muestra los valores ordenados de los factores, las variables de medidas repetidas, los sujetos de medidas repetidas y los sujetos de los efectos aleatorios junto con las frecuencias correspondientes.

**Estadísticos del modelo.** Genera tablas correspondientes a:

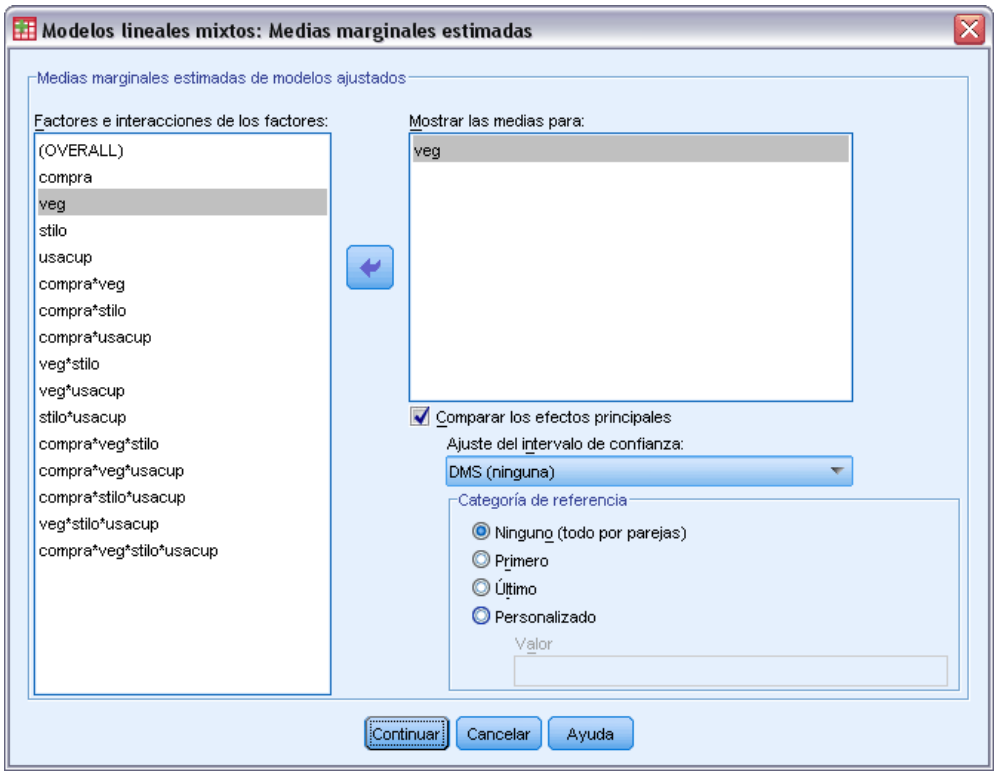
- **Estimaciones de los parámetros.** Muestra las estimaciones de los parámetros de los efectos fijos y aleatorios y los errores típicos aproximados correspondientes.
- **Contrastes sobre parámetros de covarianza.** Muestra los errores típicos asintóticos y las pruebas de Wald de los parámetros de covarianza.

- **Correlaciones de las estimaciones de los parámetros.** Muestra la matriz de correlaciones asintóticas de las estimaciones de los parámetros de los efectos fijos.
- **Covarianzas de las estimaciones de los parámetros.** Muestra la matriz de covarianzas asintóticas de las estimaciones de los parámetros de los efectos fijos.
- **Covarianzas de los efectos aleatorios.** Muestra la matriz de covarianzas estimada de los efectos aleatorios. Esta opción está disponible sólo si especifica al menos un efecto aleatorio. Si se especifica una variable de sujetos para un efecto aleatorio, se muestra el bloque común.
- **Covarianzas de los residuos.** Muestra la matriz de covarianzas residual estimada. Esta opción está disponible sólo en caso de que se haya especificado una variable para repetidas. Si se especifica una variable de sujetos, se muestra el bloque común.
- **Matriz de coeficientes del contraste.** Esta opción muestra las funciones estimables utilizadas para contrastar los efectos fijos y las hipótesis personalizadas.

**Intervalo de confianza.** Este valor se usa siempre que se genera un intervalo de confianza. Especifique un valor mayor o igual a 0 e inferior a 100. El valor por defecto es 95.

# Medias marginales estimadas de modelos lineales mixtos

Figura 5-7
*Cuadro de diálogo Medias marginales estimadas de modelos lineales mixtos*

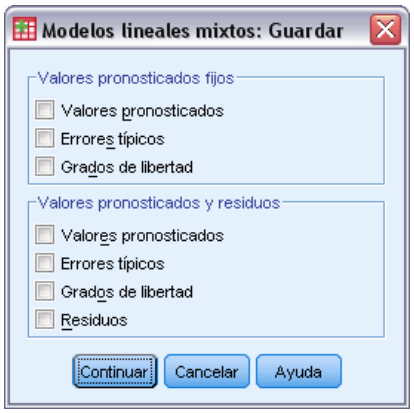


**Medias marginales estimadas de modelos ajustados.** Este grupo permite solicitar las medias marginales estimadas pronosticadas por el modelo de la variable dependiente en las casillas, así como los errores típicos correspondientes a los factores especificados. Además, puede solicitar una comparación de los niveles de los factores de los efectos principales.

- **Factores e interacciones de los factores.** La lista contiene los factores y las interacciones de los factores que se han especificado en el cuadro de diálogo Fijo, además de un término `GLOBAL`. Los términos del modelo construidos a partir de covariables no se incluyen en esta lista.
- **Mostrar las medias para.** El procedimiento calculará las medias marginales estimadas para los factores y las iteraciones de factores seleccionadas en esta lista. Si se ha seleccionado `GLOBAL`, se mostrarán las medias marginales estimadas de la variable dependiente, contrayendo todos los factores. Tenga en cuenta que permanecerán seleccionados todos los factores y las iteraciones de los factores, a menos que se haya eliminado una variable asociada de la lista Factores en el cuadro de diálogo principal.
- **Comparar los efectos principales.** Esta opción permite solicitar comparaciones por parejas de los niveles de los efectos principales seleccionados. La opción Corrección del intervalo de confianza permite aplicar ajustes a los intervalos de confianza y los valores de significación para explicar comparaciones múltiples. Los métodos disponibles son: LSD (ningún ajuste), Bonferroni y Sidak. Por último, para cada factor, se puede seleccionar la categoría de referencia con la que se realizan las comparaciones. Si no se selecciona ninguna categoría de referencia, se construirán todas las comparaciones por parejas. Las opciones disponibles para la categoría de referencia son la primera, la última o una personalizada (en cuyo caso, se introduce el valor de la categoría de referencia).

## Guardar Modelos lineales mixtos

Figura 5-8
*Cuadro de diálogo Guardar modelos lineales mixtos*




Este cuadro de diálogo le permite guardar diversos resultados del modelo en el archivo de trabajo.

**Valores pronosticados fijos.** Guarda las variables relacionadas con las medias de regresión sin los efectos.

- **Valores pronosticados.** Las medias de regresión sin los efectos aleatorios.

- **Errores típicos.** Los errores típicos de las estimaciones.
- **Grados de libertad.** Los grados de libertad asociados a las estimaciones.

**Valores pronosticados y residuos.** Guarda las variables relacionadas con el valor ajustado por el modelo.

- **Valores pronosticados.** El valor ajustado por el modelo.
- **Errores típicos.** Los errores típicos de las estimaciones.
- **Grados de libertad.** Los grados de libertad asociados a las estimaciones.
- **Residuos.** El valor de los datos menos el valor pronosticado.

## Funciones adicionales del comando MIXED

Con el lenguaje de sintaxis de comandos también podrá:

- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando `TEST`).
- Incluir los valores perdidos definidos por el usuario (utilizando el subcomando `MISSING`).
- Calcular las medias marginales estimadas de los valores especificados de las covariables (utilizando la palabra clave `WITH` del subcomando `EMMEANS`).
- Comparar los efectos principales simples de las iteraciones (utilizando el subcomando `EMMEANS`).

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

Capítulo
6

# Modelos lineales generalizados

El modelo lineal generalizado amplía el modelo lineal general, de manera que la variable dependiente está relacionada linealmente con los factores y las covariables mediante una determinada función de enlace. Además, el modelo permite que la variable dependiente tenga una distribución que no sea normal. El modelo lineal generalizado cubre los modelos estadísticos más utilizados, como la regresión lineal para las respuestas distribuidas normalmente, modelos logísticos para datos binarios, modelos loglineales para datos de recuento, modelos log-log complementario para datos de supervivencia censurados por intervalos, además de muchos otros modelos estadísticos a través de la propia formulación general del modelo.

**Ejemplos.** Una compañía de transporte puede utilizar modelos lineales generalizados para ajustar una regresión de Poisson a las frecuencias de daños de varios tipos de barcos construidos en varios períodos de tiempo. El modelo resultante puede ayudar a determinar cuales son los tipos de barcos más propensos a sufrir daños.

Una compañía de seguros de automóviles puede utilizar modelos lineales generalizados para ajustar una regresión gamma a las reclamaciones por daños de los automóviles. El modelo resultante puede ayudar a determinar los factores que más contribuyen al tamaño de la reclamación.

Los investigadores médicos pueden utilizar modelos lineales generalizados para ajustar una regresión log-log complementario a los datos de supervivencia censurados por intervalos para pronosticar el tiempo que tardará en reaparecer una enfermedad.

**Datos.** La respuesta puede ser de escala, de recuentos, binaria o eventos en ensayos. Se supone que los factores son categóricos. Las covariables, el peso de escala y el desplazamiento se suponen que son de escala.

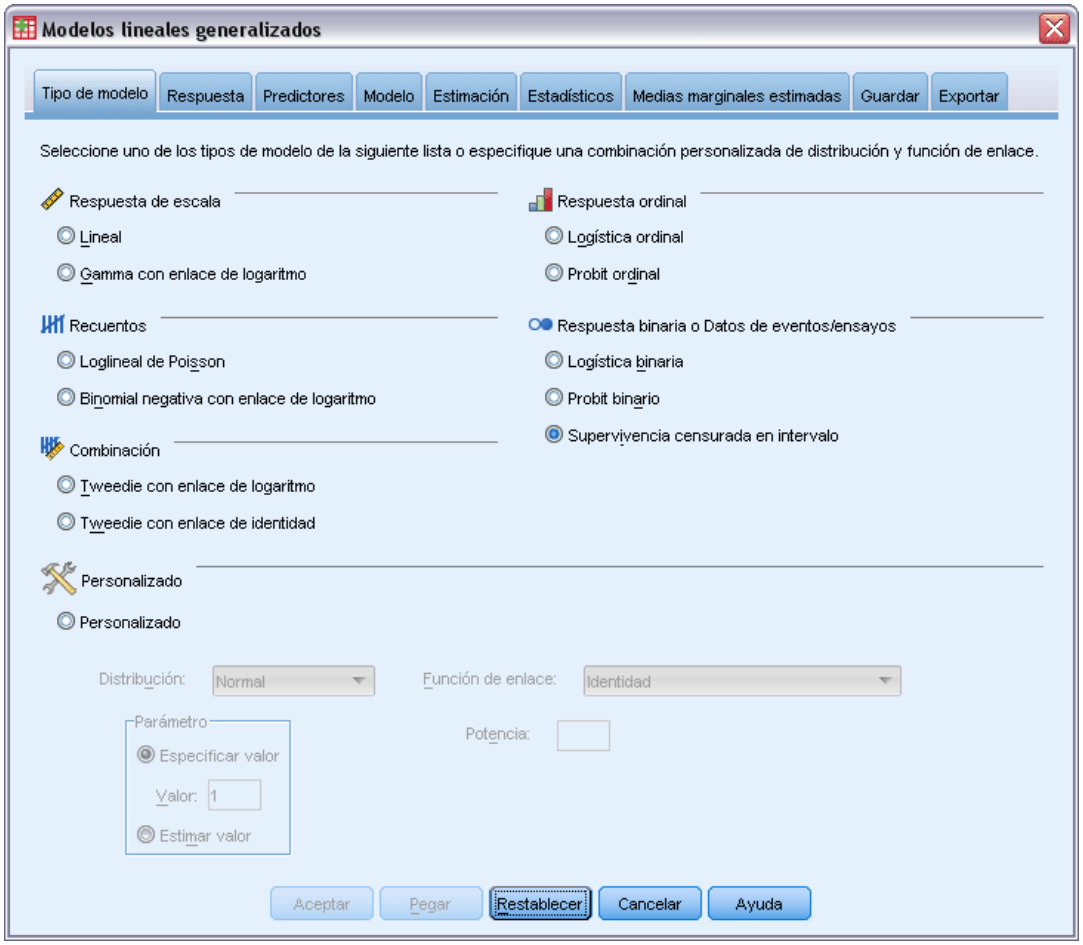
**Supuestos.** Se supone que los casos son observaciones independientes.

### *Para obtener un modelo lineal generalizado*

Seleccione en los menús:

Analizar > Modelos lineales generalizados > Modelos lineales generalizados...

Figura 6-1
*Modelos lineales generalizados: pestaña Tipo de modelo*




- Especifique una distribución y una función de enlace (consulte a continuación detalles sobre las opciones disponibles).
- En la pestaña Respuesta, seleccione una variable dependiente.
- En la pestaña Predictores, seleccione los factores y las covariables que utilizará para pronosticar la variable dependiente.
- En la pestaña Modelo, especifique los efectos del modelo utilizando las covariables y los factores seleccionados.

La pestaña Tipo de modelo permite especificar la distribución y la función de enlace del modelo, además de proporcionar accesos directos a varios modelos habituales que aparecen clasificados por tipo de respuesta.

### *Tipos de modelos*

**Respuesta de escala.**

- **Lineal.** Especifica la distribución normal y la función de enlace identidad.
- **Gamma con enlace de logaritmo.** Especifica la distribución gamma y la función de enlace de logaritmo.

**Respuesta ordinal.**

- **Logística ordinal.** Especifica la distribución multinomial (ordinal) y la función de enlace logit acumulado.
- **Probit ordinal.** Especifica la distribución multinomial (ordinal) y la función de enlace probit acumulado.

**Recuentos.**

- **Loglineal de Poisson.** Especifica la distribución de Poisson y la función de enlace de logaritmo.
- **Binomial negativa con enlace de logaritmo.** Especifica la distribución binomial negativa (con el valor 1 para el parámetro auxiliar) y la función de enlace de logaritmo. Para que el procedimiento calcule el valor del parámetro auxiliar, especifique un modelo personalizado con distribución binomial negativa y seleccione Estimar valor en el grupo de parámetros.

**Respuesta binaria o Datos de eventos/ensayos.**

- **Logística binaria.** Especifica la distribución binomial y la función de enlace logit.
- **Probit binario.** Especifica la distribución binomial y la función de enlace probit.
- **Supervivencia censurada en intervalo.** Especifica la distribución binomial y la función de enlace log-log complementario.

**Combinación.**

- **Tweedie con enlace de logaritmo.** Especifica la distribución de Tweedie y la función de enlace de logaritmo.
- **Tweedie con enlace de identidad.** Especifica la distribución de Tweedie y la función de enlace identidad.

**Personalizado.** Especifique su propia combinación de distribución y función de enlace.

### *Distribución*

Esta selección especifica la distribución de la variable dependiente. La posibilidad de especificar una distribución que no sea la normal y una función de enlace que no sea la identidad es la principal mejora que aporta el modelo lineal generalizado respecto al modelo lineal general. Hay muchas combinaciones posibles de distribución y función de enlace, varias de las cuales pueden ser adecuadas para un determinado conjunto de datos, por lo que su elección puede estar guiada por consideraciones teóricas a priori y por las combinaciones que parezcan funcionar mejor.

- **Binomial.** Esta distribución es adecuada únicamente para las variables que representan una respuesta binaria o un número de eventos.
- **Gamma.** Esta distribución es adecuada para las variables con valores de escala positivos que se desvían hacia valores positivos más grandes. Si un valor de datos es menor o igual que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.

- **De Gauss inversa.** Esta distribución es adecuada para las variables con valores de escala positivos que se desvían hacia valores positivos más grandes. Si un valor de datos es menor o igual que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **Binomial negativa.** Esta distribución considera el número de intentos necesarios para lograr *k* éxitos y es adecuada para variables que tengan valores enteros que no sean negativos. Si un valor de datos no es entero, es menor que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis. El valor del parámetro auxiliar de la distribución binomial negativa puede ser cualquier número mayor o igual que 0; se puede establecer en un valor fijo o dejar que lo estime el procedimiento. Cuando el parámetro auxiliar se establece en 0, utilizar esta distribución equivale a utilizar la distribución de Poisson.
- **Normal.** Es adecuada para variables de escala cuyos valores adoptan una distribución simétrica con forma de campana en torno a un valor central (la media). La variable dependiente debe ser numérica.
- **Poisson.** Esta distribución considera el número de ocurrencias de un evento de interés en un período fijo de tiempo y es apropiada para variables que tengan valores enteros que no sean negativos. Si un valor de datos no es entero, es menor que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **Tweedie.** Esta distribución es adecuada para variables que puedan representarse mediante mezclas de Poisson de distribuciones gamma; la distribución es una "mezcla" en el sentido de que combina las propiedades de distribuciones continuas (toma valores reales no negativos) y discretas (masa de probabilidad positiva en un único valor, 0). La variable dependiente debe ser numérica y los valores de los datos deben ser iguales o mayores que cero. Si un valor de datos es menor que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis. El valor fijo del parámetro de la distribución de Tweedie puede ser cualquier número mayor que uno y menor que dos.
- **Multinomial.** Esta distribución es adecuada para variables que representan una respuesta ordinal. La variable dependiente puede ser numérica o de cadena, y debe tener como mínimo dos valores válidos distintos de los datos.

### *Funciones de enlace*

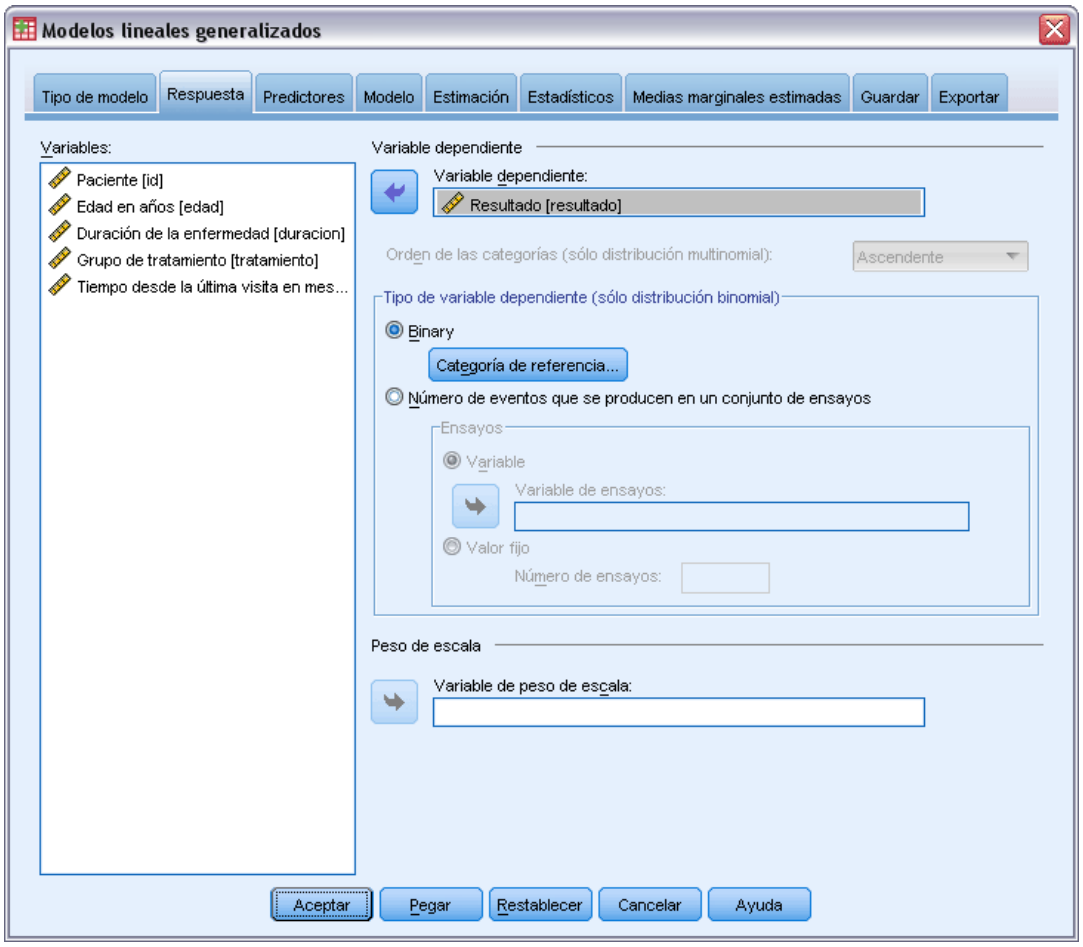
La función de enlace es una transformación de la variable dependiente que permite la estimación del modelo. Se encuentran disponibles las siguientes funciones:

- **Identidad.** $f(x)=x$. No se transforma la variable dependiente. Este vínculo se puede utilizar con cualquier distribución.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Es apropiada únicamente para la distribución binomial.
- **Cauchit acumulada.** $f(x) = \tan(\pi\ (x - 0.5))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Log-log complementario acumulado.** $f(x)=\ln(-\ln(1-x))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Logit acumulado.** $f(x)=\ln(x / (1-x))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta.. Es apropiada únicamente para la distribución multinomial.
- **Log-log negativo acumulado.** $f(x)=(-\ln(-\ln x))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.

- **Probit acumulada.** $f(x)=\Phi^{-1}(x)$, aplicada a la probabilidad acumulada de cada categoría de la respuesta, donde $\Phi^{-1}$ es la función de distribución acumulada normal típica inversa. Es apropiada únicamente para la distribución multinomial.
- **Log.** $f(x)=\log(x)$. Este vínculo se puede utilizar con cualquier distribución.
- **Complemento log.** $f(x)=\log(1-x)$. Es apropiada únicamente para la distribución binomial.
- **Logit.** $f(x)=\log(x / (1-x))$. Es apropiada únicamente para la distribución binomial.
- **Binomial negativa.** $f(x)=\log(x / (x+k^{-1}))$, donde $k$ es el parámetro auxiliar de la distribución binomial negativa. Es apropiada únicamente para la distribución binomial negativa.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Es apropiada únicamente para la distribución binomial.
- **Potencia de las ventajas.** $f(x)=[(x/(1-x))^{\alpha}-1]/\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. $\alpha$ es la especificación de número necesaria y debe ser un número real. Es apropiada únicamente para la distribución binomial.
- **Probit.** $f(x)=\Phi^{-1}(x)$, donde $\Phi^{-1}$ es la función de distribución acumulada normal típica inversa. Es apropiada únicamente para la distribución binomial.
- **Potencia.** $f(x)=x^{\alpha}$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. $\alpha$ es la especificación de número necesaria y debe ser un número real. Este vínculo se puede utilizar con cualquier distribución.

# Modelos lineales generalizados: Respuesta

Figura 6-2
*Cuadro de diálogo Modelos lineales generalizados*




En muchos casos, puede especificar sencillamente una variable dependiente. No obstante, las variables que adoptan únicamente dos valores y las respuestas que registran eventos en ensayos que requieren una atención adicional.

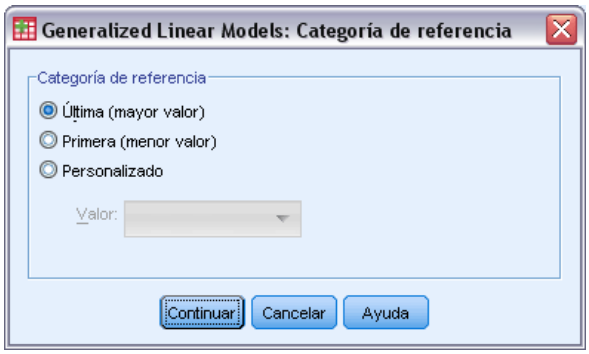
- **Respuesta binaria.** Cuando la variable dependiente adopta únicamente dos valores, puede especificar la categoría de referencia para la estimación de los parámetros. Una variable de respuesta binaria puede ser de cadena o numérica.
- **Número de eventos que se producen en un conjunto de ensayos** Cuando la respuesta es un número de eventos que ocurren en un conjunto de ensayos, la variable dependiente contiene el número de eventos y puede seleccionar una variable adicional que contenga el número de ensayos. Otra posibilidad, si el número de ensayos es el mismo en todos los sujetos, consiste en especificar los ensayos mediante un valor fijo. El número de ensayos debe ser mayor o igual que el número de eventos para cada caso. Los eventos deben ser enteros no negativos y los ensayos deben ser enteros positivos.

Para los modelos multinomiales ordinales, puede especificar el orden de las categorías de la respuesta: ascendente, descendente o datos (el orden de los datos indica que el primer valor encontrado en los datos define la primera categoría y el último valor encontrado define la última categoría).

**Peso de escala.** El parámetro de escala es un parámetro del modelo estimado relacionado con la varianza de la respuesta. Los pesos de escala son valores "conocidos" que pueden variar de una observación a otra. Si se especifica una variable de peso de escala, el parámetro de escala, que está relacionado con la varianza de la respuesta, se divide por él para cada observación. Los casos cuyo valor del peso de escala es menor o igual que 0 o que son perdidos no se utilizan en el análisis.

## *Modelos lineales generalizados: Categoría de referencia*

Figura 6-3
*Cuadro de diálogo Modelos lineales generalizados: Categoría de referencia*

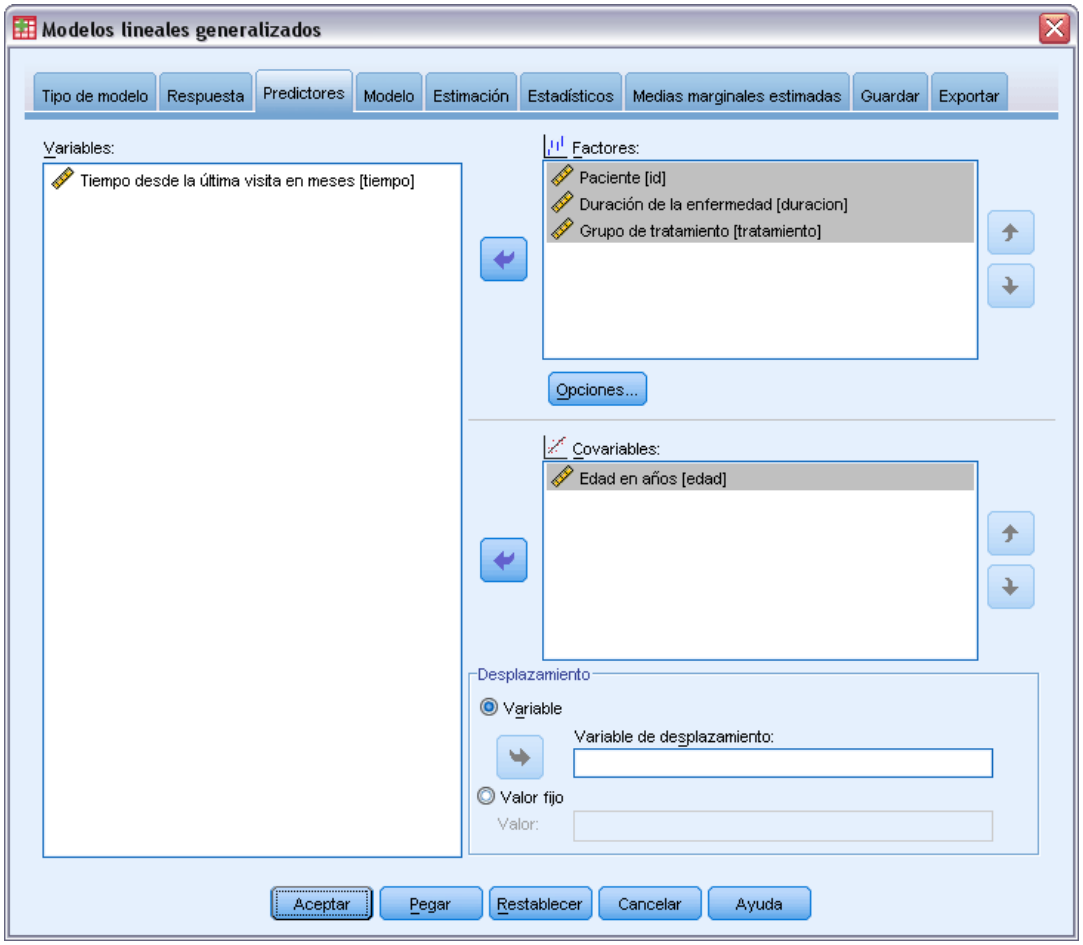



Para una respuesta binaria, puede elegir la categoría de referencia de la variable dependiente. Puede afectar a ciertos resultados, como las estimaciones de los parámetros y los valores guardados, pero no debería cambiar el ajuste del modelo. Por ejemplo, si la respuesta binaria toma los valores 0 y 1:

- Por defecto, el procedimiento utiliza la última categoría (la de mayor valor), o 1, como la categoría de referencia. En esta situación, las probabilidades guardadas por el modelo estiman la posibilidad de que un determinado caso tome el valor 0 y las estimaciones de los parámetros deben interpretarse como relativas a la probabilidad de la categoría 0.
- Si especifica la primera categoría (la de menor valor), o 0, como la categoría de referencia, las probabilidades guardadas por el modelo estimarán la posibilidad de que un determinado caso tome el valor 1.
- Si especifica la categoría personalizada y la variable tiene etiquetas definidas, puede establecer la categoría de referencia eligiendo un valor de la lista. Puede resultar cómodo si, mientras se especifica un modelo, no recuerda exactamente cómo se ha codificado una determinada variable.

# *Modelos lineales generalizados: Predictores*

Figura 6-4
*Modelos lineales generalizados: Pestaña Predictores*




La pestaña Predictores permite especificar los factores y las covariables que se utilizarán para crear los efectos del modelo y para especificar un desplazamiento opcional.

**Factores.** Los factores son predictores categóricos y pueden ser numéricos o de cadena.

**Covariables.** Las covariables son predictores de escala y deben ser numéricas.

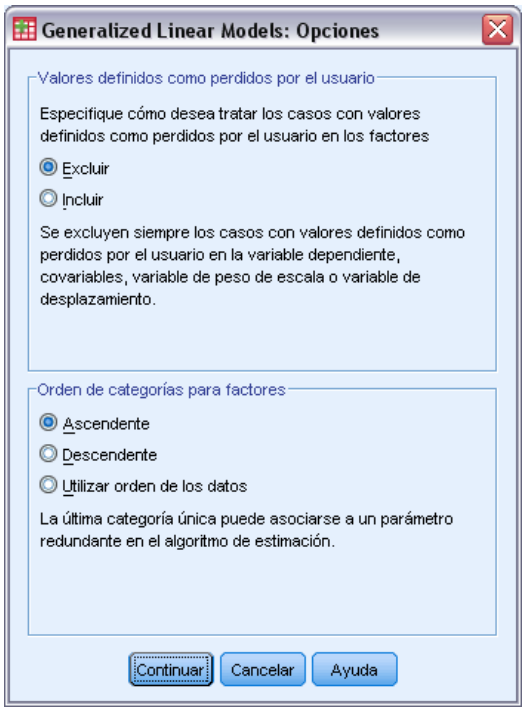
*Nota*: Cuando la respuesta es binomial con formato binario, el procedimiento calcula los estadísticos de bondad de ajuste de chi cuadrado y de desvianza por subpoblaciones que se basan en la clasificación cruzada de los valores observados de los factores y las covariables seleccionadas. Debe mantener el mismo conjunto de predictores en las diferentes ejecuciones del procedimiento para asegurarse de que se utiliza un número coherente de subpoblaciones.

**Desplazamiento.** El término desplazamiento es un predictor "estructural". Su coeficiente no se estima por el modelo pero se supone que tiene el valor 1. Por tanto, los valores del desplazamiento se suman sencillamente al predictor lineal de la variable dependiente. Esto resulta especialmente útil en los modelos de regresión de Poisson, en los que cada caso pueden tener diferentes niveles

de exposición al evento de interés. Por ejemplo, al modelar las tasas de accidente de diferentes conductores, hay una importante diferencia entre un conductor que ha sido el culpable de un accidente en tres años y un conductor que ha sido el culpable de un accidente en 25 años. El número de accidentes se puede modelar como una respuesta de Poisson si la experiencia del conductor se incluye como un término de desplazamiento.

## Modelos lineales generalizados: Opciones

Figura 6-5
*Cuadro de diálogo Modelos lineales generalizados: Opciones*



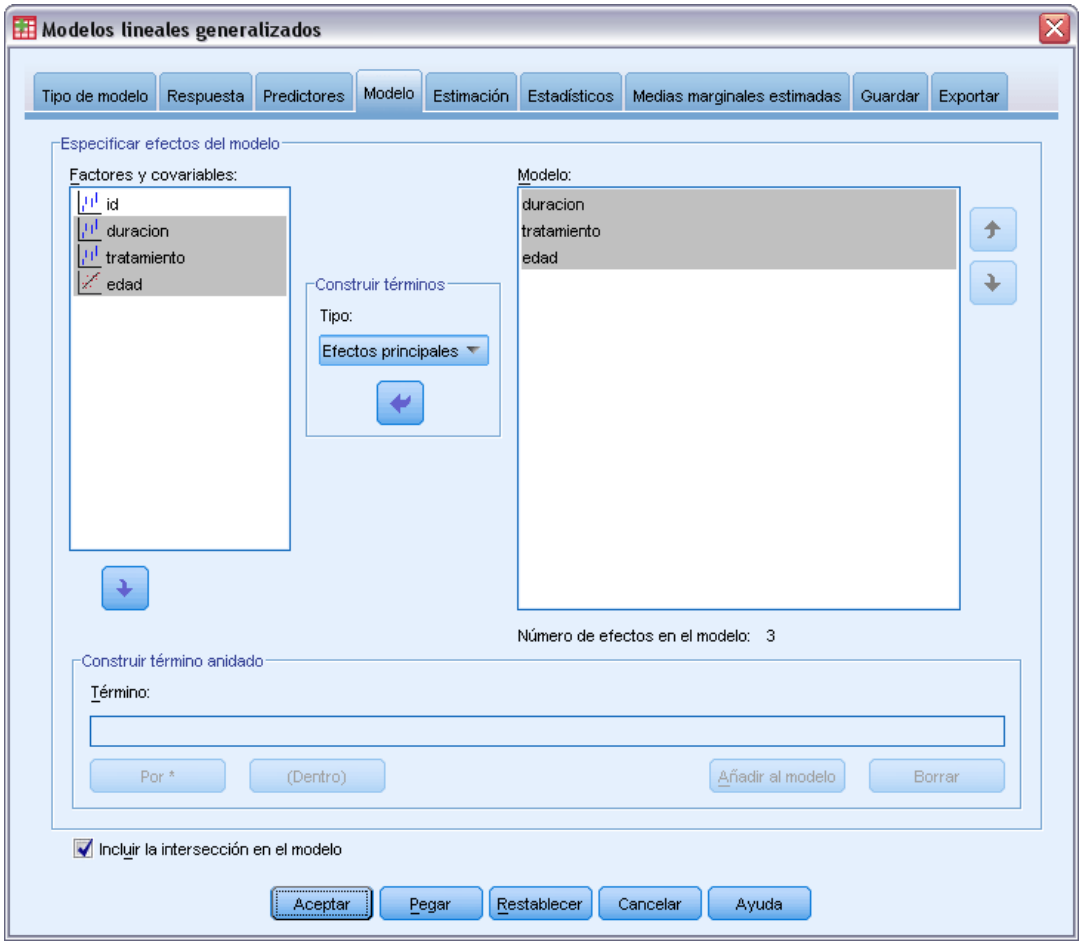

Estas opciones se aplican a todos los factores especificados en la pestaña Predictores.

**Valores definidos como perdidos por el usuario.** Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores definidos como perdidos por el usuario se deben tratar como válidos entre las variables de factor.

**Orden de categorías.** Es relevante para determinar el último nivel de un factor, que puede estar asociado a un parámetro redundante del algoritmo de estimación. Si se cambia el orden de categorías es posible que cambien también los valores de los efectos de los niveles de los factores, ya que estas estimaciones de los parámetros se calculan respecto al "último" nivel. Los factores se pueden ordenar en orden ascendente desde el valor mínimo hasta el máximo, en orden descendente desde el valor máximo hasta el mínimo o siguiendo el "orden de los datos". Significa que el primer valor encontrado en los datos define la primera categoría, y el último valor único encontrado define la última categoría.

# *Modelos lineales generalizados: Modelo*

Figura 6-6
*Modelos lineales generalizados: Pestaña Modelo*




**Especificar efectos del modelo.** El modelo por defecto sólo utiliza la intersección, por lo que deberá especificar explícitamente todos los demás efectos del modelo. Puede elegir entre términos anidados o no anidados.

### *Términos no anidados*

Para las covariables y los factores seleccionados:

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

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

**Factorial.** Crea todas las interacciones y efectos principales posibles para las variables seleccionadas.

**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.

### *Términos anidados*

En este procedimiento, puede construir términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción, como términos polinómicos que implican a la misma covariable, o añadir varios niveles de anidación al término anidado.

**Limitaciones.** Existen las siguientes restricciones para los términos anidados:

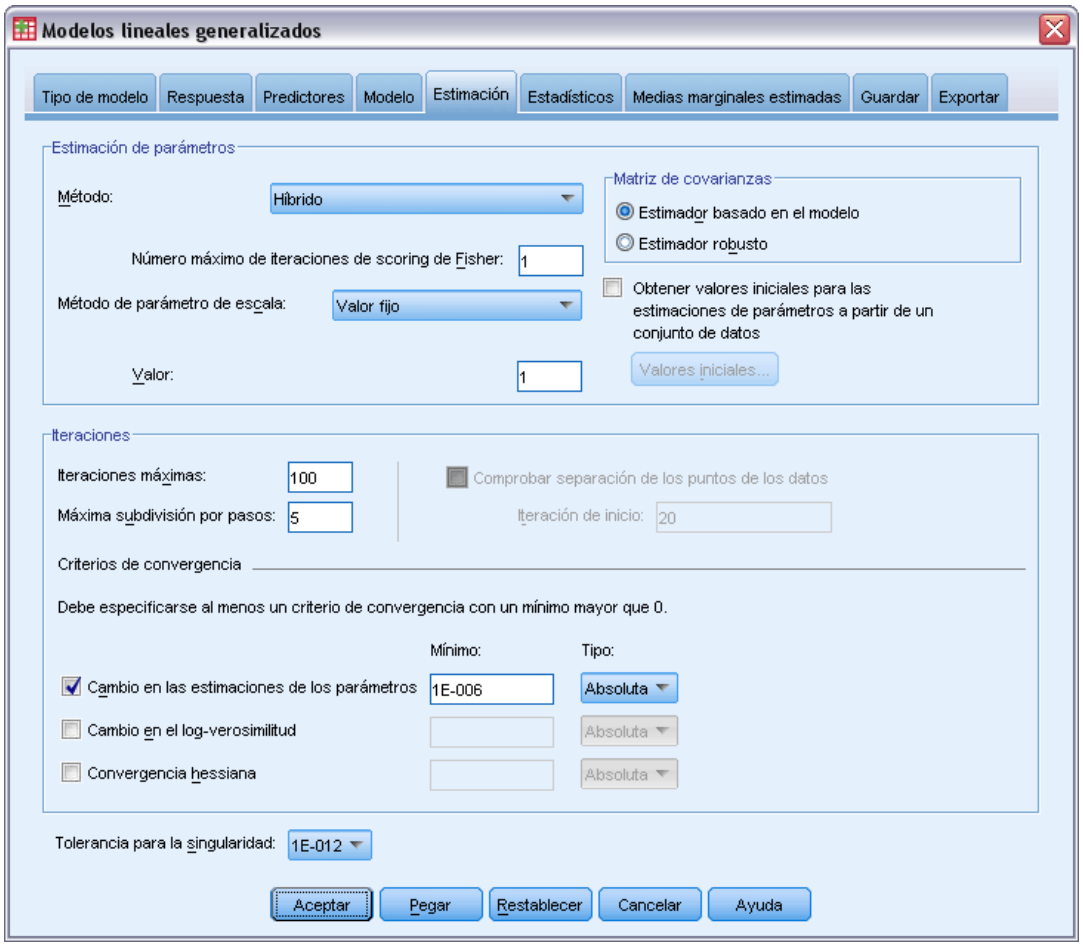
- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A*A*.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A(A)*.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si *A* es un factor y *X* es una covariable, no es válido especificar *A(X)*.

**Intersección.** La intersección se incluye normalmente en el modelo. Si asume que los datos pasan por el origen, puede excluir la intersección.

Los modelos con distribución ordinal multinomial no tienen un único término de intersección, sino que tienen parámetros de umbral que definen los puntos de transición entre las categorías adyacentes. Los umbrales siempre se incluyen en el modelo.

# Modelos lineales generalizados: Estimación

Figura 6-7
*Modelos lineales generalizados: Pestaña Estimación*




**Estimación de parámetros.** Los controles de este grupo le permiten especificar los métodos de estimación y proporcionar los valores iniciales para las estimaciones de los parámetros.

- **Método.** Puede seleccionar el método de estimación de los parámetros. Los métodos disponibles son Newton-Raphson, Scoring de Fisher o un método híbrido en el que las iteraciones de Scoring de Fisher se realizan antes de cambiar al método de Newton-Raphson. Si se logra la convergencia durante la fase de Scoring de Fisher del método híbrido antes de que se lleven a cabo el número máximo de iteraciones de Fisher, el algoritmo continúa con el método de Newton-Raphson.
- **Método de parámetro de escala.** Puede seleccionar el método de estimación del parámetro de escala. La máxima verosimilitud estima conjuntamente el parámetro de escala y los efectos del modelo. Tenga en cuenta que esta opción no es válida si la respuesta sigue una distribución binomial negativa, de Poisson, binomial o multinomial. Las opciones de desvianza y de chi-cuadrado de Pearson estiman el parámetro de escala a partir del valor de dichos estadísticos. Otra posibilidad consiste en especificar un valor corregido para el parámetro de escala.

- **Valores iniciales.** El procedimiento calculará automáticamente los valores iniciales de los parámetros. Otra posibilidad consiste en especificar los valores iniciales de las estimaciones de los parámetros.
- **Matriz de covarianzas.** El estimador basado en el modelo es la negativa de la inversa generalizada de la matriz hessiana. El estimador robusto (también llamado el estimador de Huber/White/Sandwich) es un estimador basado en el modelo "corregido" que proporciona una estimación coherente de la covarianza, incluso cuando la varianza y las funciones de enlace no se han especificado correctamente.

**Iteraciones.**

- **Iteraciones máximas.** Número máximo de iteraciones que se ejecutará el algoritmo. Especifique un número entero no negativo.
- **Máxima subdivisión por pasos.** En cada iteración, se reduce el tamaño del paso mediante un factor de 0,5 hasta que aumenta el logaritmo de la verosimilitud o se alcanza la máxima subdivisión por pasos. Especifique un número entero positivo.
- **Comprobar si hay separación completa de los puntos de los datos.** Si se activa, el algoritmo realiza una prueba para garantizar que las estimaciones de los parámetros tienen valores exclusivos. Se produce una separación cuando el procedimiento pueda generar un modelo que clasifique cada caso de forma correcta. Esta opción no está disponible para respuestas multinomiales y binomiales con formato binario.

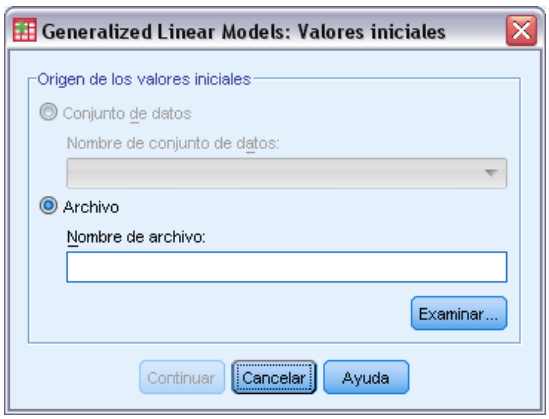
**Criterios de convergencia.**

- **Convergencia de los parámetros.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en las estimaciones de los parámetros son inferiores al valor especificado, que debe ser positivo.
- **Convergencia del logaritmo de la verosimilitud.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en la función de log-verosimilitud sean inferiores que el valor especificado, que debe ser positivo.
- **Convergencia hessiana.** En el caso de la especificación Absoluta, se supone la convergencia si un estadístico basado en la convergencia hessiana es menor que el valor positivo especificado. En el caso de la especificación Relativa, se supone la convergencia si el estadístico es menor que el producto del valor positivo especificado y el valor absoluto del logaritmo de la verosimilitud.

**Tolerancia para la singularidad.** Las matrices singulares (que no se pueden invertir) tienen columnas linealmente dependientes, lo que causar graves problemas al algoritmo de estimación. Incluso las matrices casi singulares pueden generar resultados deficientes, por lo que el procedimiento tratará una matriz cuyo determinante es menor que la tolerancia como singular. Especifique un valor positivo.

## Modelos lineales generalizados: Valores iniciales

Figura 6-8
*Cuadro de diálogo Modelos lineales generalizados: Valores iniciales*




Si se especifican valores iniciales, deben proporcionarse para todos los parámetros del modelo (incluidos los parámetros redundantes). En el conjunto de datos, el orden de las variables de izquierda a derecha debe ser: *TipoFila_*, *NombreVar_*, *P1*, *P2*, …, donde *TipoFila_* y *NombreVar_* son variables de cadena y *P1*, *P2*, … son variables numéricas que corresponden a una lista ordenada de los parámetros.

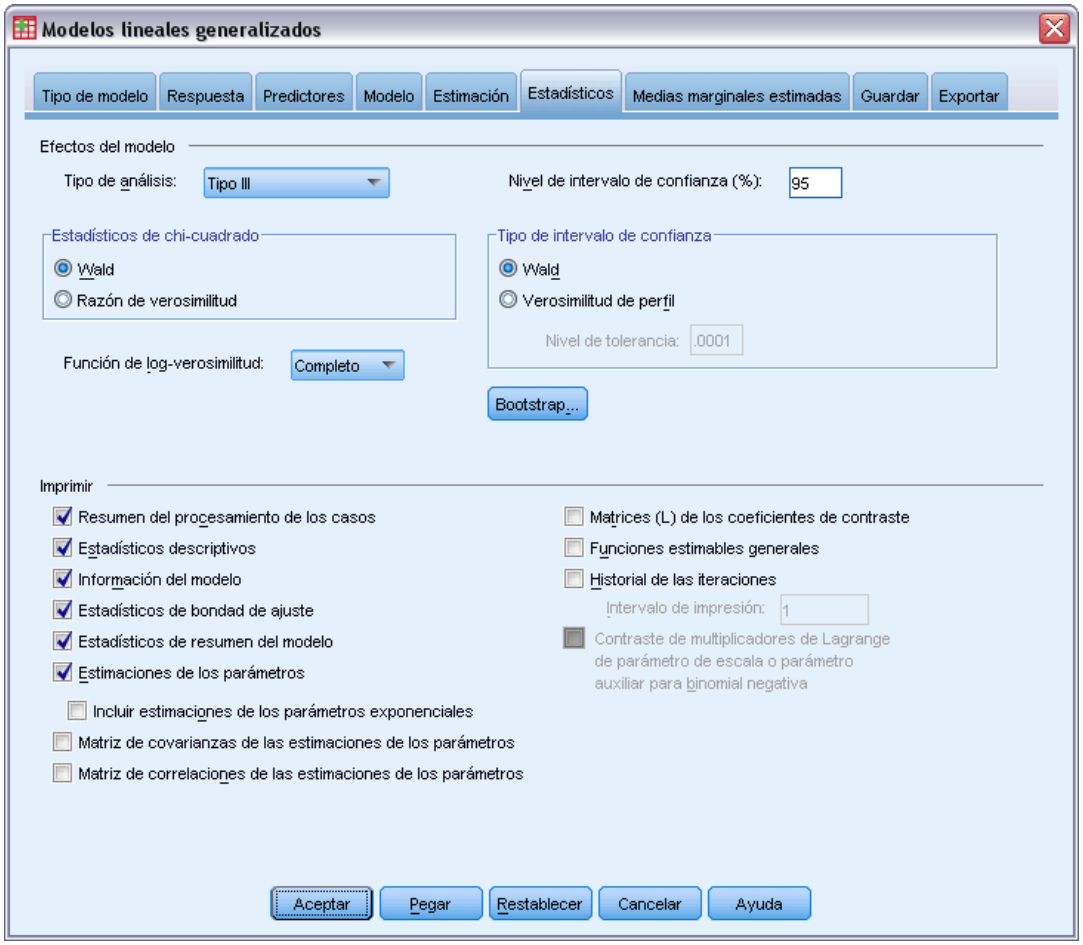
- Los valores iniciales se proporcionan en un registro con el valor *EST* para la variable *TipoFila_*; los valores iniciales reales se proporcionan en las variables *P1*, *P2*, …. El procedimiento ignora todos los registros para los que *TipoFila_* tienen un valor diferente de *EST*, así como todos los registros posteriores a la primera aparición de *TipoFila_* igual a *EST*.
- La intersección, si se incluye en el modelo, o los parámetros de umbral, si la respuesta sigue una distribución multinomial, deben ser los primeros valores iniciales.
- El parámetro de escala y, si la respuesta sigue una distribución binomial negativa, el parámetro binomial negativo, deben ser los últimos valores iniciales especificados.
- Si está activo Segmentar archivo, las variables deberán comenzar con la variable (o las variables) de segmentación del archivo en el orden especificado al crear la segmentación del archivo, seguidas de *TipoFila_*, *NombreVar_*, *P1*, *P2*, … como se ha indicado anteriormente. La segmentación debe haberse realizado en el conjunto de datos especificado en el mismo orden que en el conjunto de datos original.

*Nota*: Los nombres de las variables *P1*, *P2*, … no son necesarios. El procedimiento aceptará cualquier nombre de variable válido para los parámetros, ya que la asignación de las variables a los parámetros se basa en la posición de la variable y no en el nombre de la variable. Se ignorarán todas las variables que aparezcan después del último parámetro.

La estructura de archivo de los valores iniciales es la misma que la utilizada al exportar el modelo como datos. Por tanto, puede utilizar los valores finales de una ejecución del procedimiento como entrada de una ejecución posterior.

# Modelos lineales generalizados: Estadísticos

Figura 6-9
*Modelos lineales generalizados: Pestaña Estadísticos*




**Efectos del modelo.**

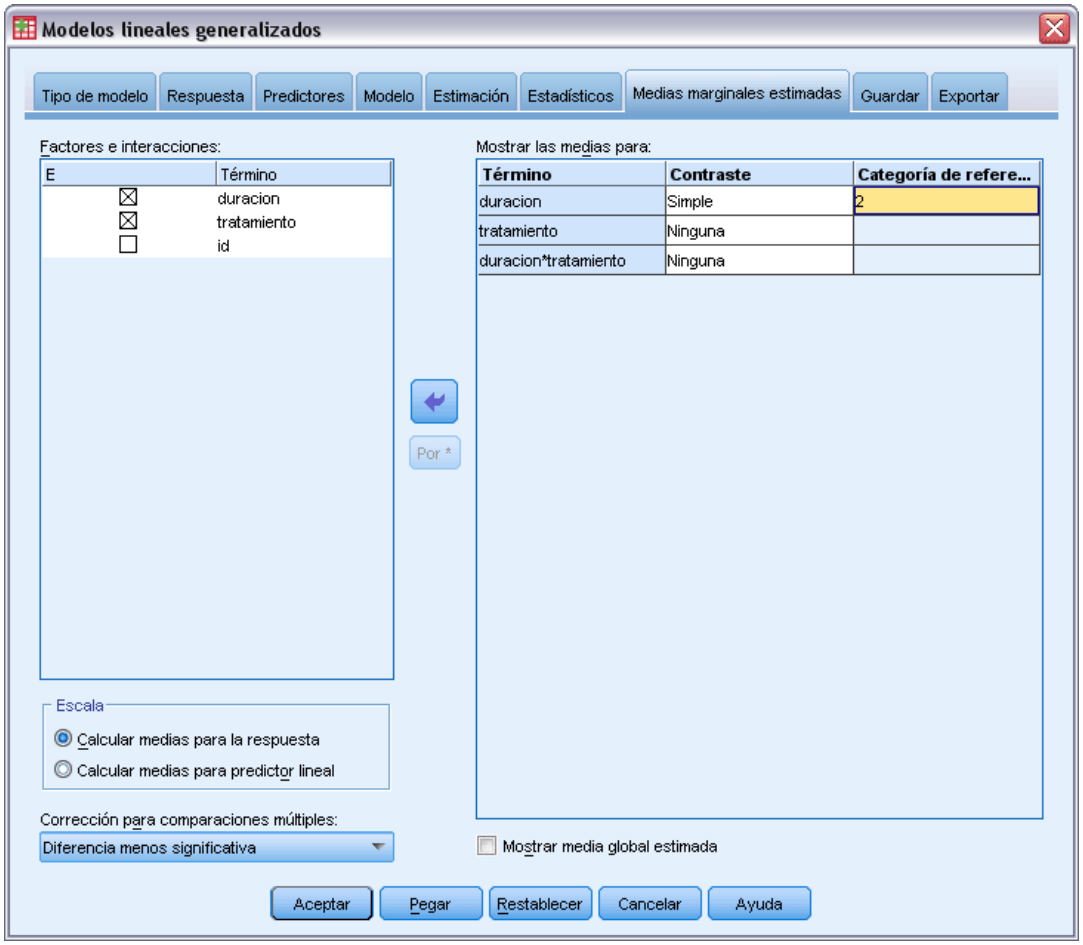
- **Tipo de análisis** Especifique el tipo de análisis que desea generar. El análisis de tipo I suele ser apropiado cuando tiene motivos a priori para ordenar los predictores del modelo, mientras que el tipo III es de aplicación más general. Los estadísticos de razón de verosimilitud o de Wald se calculan a partir de la selección realizada en el grupo Estadísticos de chi-cuadrado.
- **Intervalos de confianza.** Especifique un nivel de confianza mayor que 50 y menor que 100. Los intervalos de Wald se basan en el supuesto de que los parámetros siguen una distribución normal asintótica; los intervalos de verosimilitud de perfil son más precisos pero es posible que también requieran bastantes recursos informáticos. El nivel de tolerancia de los intervalos de verosimilitud de perfil es el criterio utilizado para detener el algoritmo iterativo utilizado para calcular los intervalos.
- **Función de log-verosimilitud.** Controla el formato de presentación de la función de log-verosimilitud. La función completa incluye un término adicional que es constante respecto a las estimaciones de los parámetros. No tiene ningún efecto sobre la estimación de los parámetros y se deja fuera de la presentación en algunos productos de software.

**Imprimir.**Los siguientes resultados están disponibles:

- **Resumen del procesamiento de los casos.** Muestra el número y el porcentaje de los casos incluidos y excluidos del análisis y la tabla Resumen de datos correlacionados.
- **Estadísticos descriptivos.** Muestra estadísticos descriptivos e información resumida acerca de los factores, las covariables y la variable dependiente.
- **Información del modelo.** Muestra el nombre del conjunto de datos, la variable dependiente o las variables de eventos y ensayos, la variable de desplazamiento, la distribución de probabilidad y la función de enlace.
- **Estadísticos de bondad de ajuste.** Muestra la desvianza y la desvianza escala, chi cuadrado de Pearson y chi cuadrado de Pearson escalado, el log-verosimilitud, el criterio de información de Akaike (AIC), AIC corregido para muestras finitas (AICC), criterio de información bayesiano (BIC), AIC consistente (CAIC).
- **Estadísticos de resumen del modelo.** Muestra contraste de ajuste del modelo, incluidos los estadísticos de la razón de la verosimilitud para el contraste ómnibus del ajuste del modelo y los estadísticos para los contrastes de tipo I o III para cada efecto.
- **Estimaciones de los parámetros.** Muestra las estimaciones de los parámetros y los correspondientes estadísticos de contraste e intervalos de confianza. Si lo desea, puede mostrar las estimaciones exponenciadas de los parámetros además de las estimaciones brutas de los parámetros.
- **Matriz de covarianzas de las estimaciones de los parámetros.** Muestra la matriz de covarianzas de los parámetros estimados.
- **Matriz de correlaciones de las estimaciones de los parámetros.** Muestra la matriz de correlaciones de los parámetros estimados.
- **Matrices (L) de los coeficientes de contraste.** Muestra los coeficientes de los contrastes para los efectos por defecto y para las medias marginales estimadas, si se solicitaron en la pestaña Medias marginales estimadas.
- **Funciones estimables generales.** Muestra las matrices para generar las matrices (L) de los coeficientes de contraste.
- **Historial de iteraciones.** Muestra el historial de iteraciones de las estimaciones de los parámetros y la log-verosimilitud, e imprime la última evaluación del vector de gradiente y la matriz hessiana. La tabla del historial de iteraciones muestra las estimaciones de los parámetros para cada *n* iteraciones a partir de la iteración 0 (las estimaciones iniciales), donde *n* es el valor del intervalo de impresión. Si se solicita el historial de iteraciones, la última iteración siempre se muestra independientemente de *n*.
- **Contraste de multiplicador de Lagrange.** Muestra los estadísticos de contraste de multiplicador de Lagrange para evaluar la validez de un parámetro de escala que se calcula utilizando la desvianza o el chi-cuadrado de Pearson, o se establece en un número fijo para las distribuciones normal, gamma y de Gauss inversa y Tweedie. Para la distribución binomial negativa, sirve como contraste del parámetro auxiliar fijo.

# Modelos lineales generalizados: Medias marginales estimadas

Figura 6-10
*Modelos lineales generalizados: Pestaña Medias marginales estimadas*




Esta pestaña permite ver medias marginales estimadas para los niveles de factores y las interacciones de los factores. También se puede solicitar que se muestre la media estimada global. Las medias marginales estimadas no están disponibles para modelos multinomiales ordinales.

**Factores e interacciones.** Esta lista contiene los factores especificados en la pestaña Predictores y las interacciones de los factores especificadas en la pestaña Modelo. Las covariables se excluyen de esta lista. Los términos pueden seleccionar directamente en esta lista o combinarse en un término de interacción utilizando el botón Por *.

**Mostrar las medias para.** Se calculan las medias estimadas de los factores seleccionados y las interacciones de los factores. El contraste determina como se configuran los contrastes de hipótesis para comparar las medias estimadas. El contraste simple requiere una categoría de referencia o un nivel de factor con el que comparar los demás.

- **Por parejas.** Se calculan las comparaciones por parejas para todas las combinaciones de niveles de los factores especificados o implicados. Este contraste es el único disponible para las interacciones de los factores.
- **Simple.** Compara la media de cada nivel con la media de un nivel especificado. Este tipo de contraste resulta útil cuando existe un grupo de control.
- **Desviación.** Cada nivel del factor se compara con la media global. Los contrastes de desviación no son ortogonales.
- **Diferencia.** Compara la media de cada nivel (excepto el primero) con la media de los niveles anteriores. En ocasiones se les denomina contrastes de Helmert invertidos.
- **Helmert.** Compara la media de cada nivel del factor (excepto el último) con la media de los niveles siguientes.
- **Repetido.** Compara la media de cada nivel (excepto el último) con la media del nivel siguiente.
- **Polinómico.** Compara el efecto lineal, cuadrático, cúbico, etc. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, y así sucesivamente. Estos contrastes se utilizan a menudo para estimar las tendencias polinómicas.

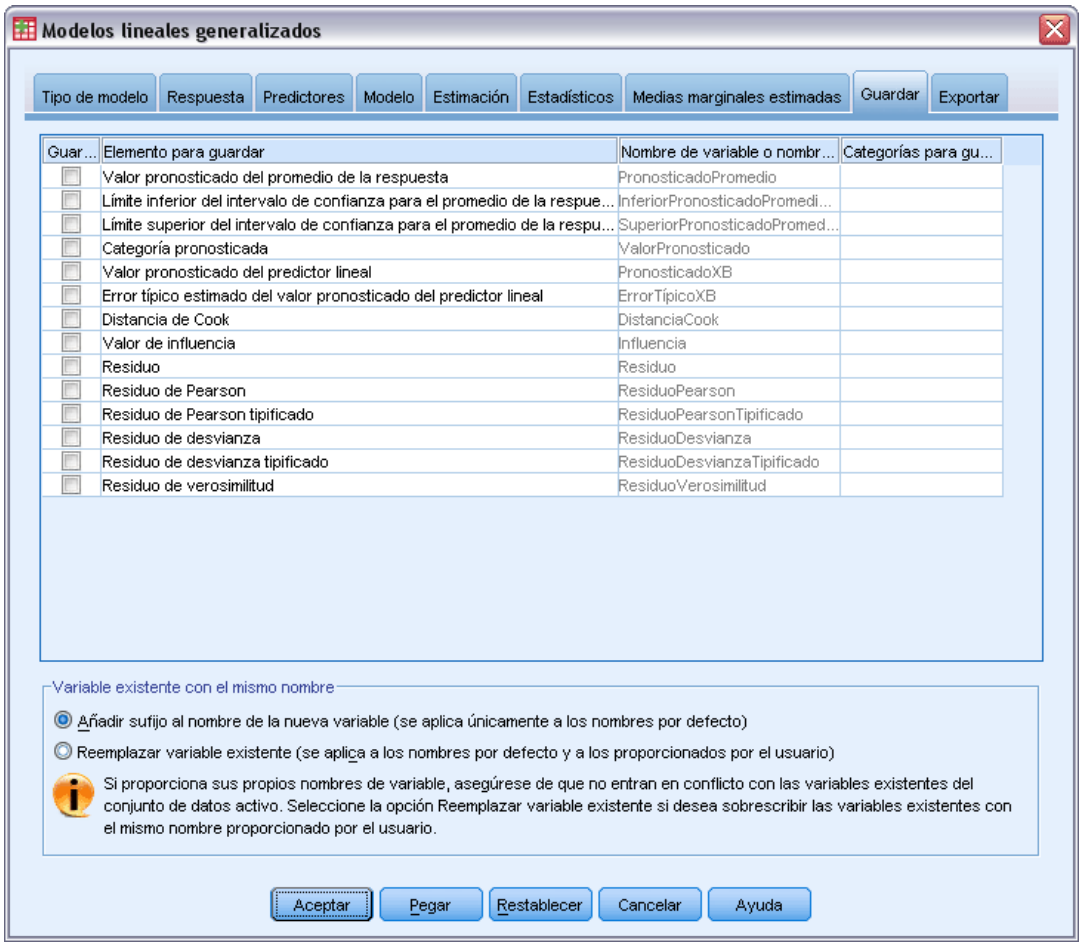
**Escala.** Se pueden calcular las medias marginales estimadas de la respuesta, basadas en la escala original de la variable dependiente o, para el predictor lineal, basadas en la variable dependiente tal como la transforma la función de enlace.

**Corrección para comparaciones múltiples.** Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Este grupo permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- **Bonferroni.** Este método corrige el nivel de significación observado por el hecho de que se están poniendo a prueba múltiples contrastes.
- **Bonferroni secuencial.** Este es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- **Sidak.** Este método ofrece límites más estrechos que los de la aproximación de Bonferroni.
- **Sidak secuencial.** Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.

# Modelos lineales generalizados: Guardar

Figura 6-11
*Modelos lineales generalizados: Pestaña Guardar*




Los elementos marcados se guardan con el nombre especificado. Puede elegir si desea sobrescribir las variables existentes con el mismo nombre que las nuevas variables o evitar conflictos de nombres adjuntando sufijos para asegurarse de que los nombres de las nuevas variables son únicos.

- **Valor pronosticado del promedio de la respuesta.** Guarda los valores pronosticados por el modelo para cada caso en la métrica de respuesta original. Cuando la distribución de la respuesta es binomial y la variable dependiente es binaria, el procedimiento guarda las probabilidades pronosticadas. Cuando la distribución de la respuesta es multinomial, la etiqueta del elemento se convierte en Probabilidad pronosticada acumulada y el procedimiento guarda la probabilidad pronosticada acumulada de cada categoría de la respuesta, salvo la última, hasta el número de categorías que se ha especificado que se guarden.
- **Límite inferior del intervalo de confianza para el promedio de la respuesta.** Guarda el límite inferior del intervalo de confianza de la media de la respuesta. Cuando la distribución de la respuesta es multinomial, la etiqueta del elemento se convierte en Límite inferior del intervalo de confianza para la probabilidad pronosticada acumulada y el procedimiento guarda el límite

inferior de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.

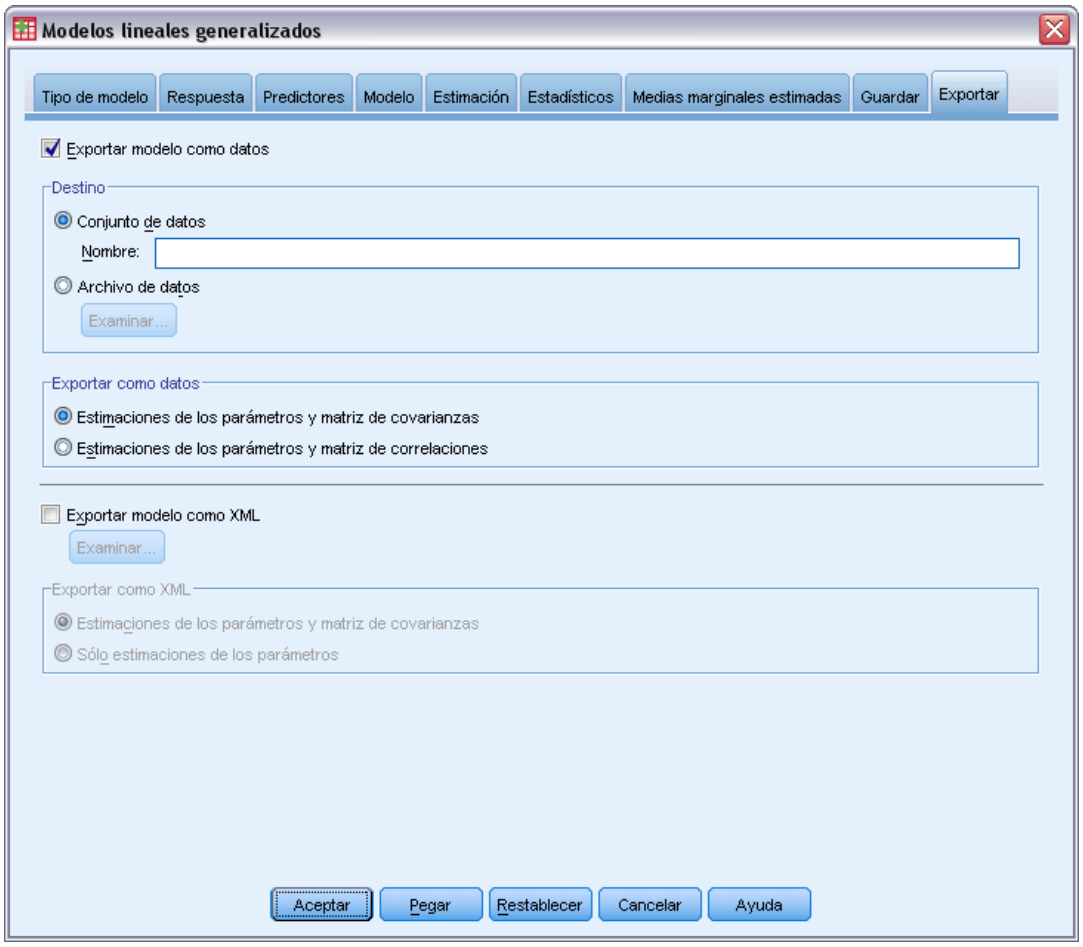
- **Límite superior del intervalo de confianza para el promedio de la respuesta.** Guarda el límite superior del intervalo de confianza de la media de la respuesta. Cuando la distribución de la respuesta es multinomial, la etiqueta del elemento se convierte en Límite superior del intervalo de confianza para la probabilidad pronosticada acumulada y el procedimiento guarda el límite superior de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.
- **Categoría pronosticada.** Para los modelos con distribución binomial y variable dependiente binaria, o distribución multinomial, esta opción guarda la categoría de respuesta pronosticada para cada caso. Esta opción no esta disponible para otras distribuciones de la respuesta.
- **Valor pronosticado del predictor lineal.** Guarda los valores pronosticados por el modelo para cada caso en la métrica del predictor lineal (respuesta transformada mediante la función de enlace especificada). Cuando la distribución de respuesta es multinomial, el procedimiento guarda el valor pronosticado de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.
- **Error típico estimado del valor pronosticado del predictor lineal.** Cuando la distribución de respuesta es multinomial, el procedimiento guarda el error típico estimado de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.

Los siguientes elementos no están disponibles cuando la distribución de la respuesta es multinomial.

- **Distancia 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. Una Distancia de Cook grande indica que la exclusión de ese caso del cálculo de los estadísticos de regresión hará variar substancialmente los coeficientes.
- **Valor de influencia.** Mide la influencia de un punto en el ajuste de la regresión. Influencia centrada varía entre 0 (no influye en el ajuste) a (N-1)/N.
- **Residuo bruto.** Diferencia entre un valor observado y el valor pronosticado por el modelo.
- **Residuo de Pearson.** La raíz cuadrada de la contribución de un caso al estadístico chi-cuadrado de Pearson, con el signo del residuo bruto.
- **Residuo de Pearson tipificado.** El residuo de Pearson multiplicado por la raíz cuadrada de la inversa del producto del parámetro de escala y 1−influencia del caso.
- **Residuo de desvianza.** La raíz cuadrada de la contribución de un caso al estadístico de desvianza, con el signo del residuo bruto.
- **Residuo de desvianza tipificado.** El residuo de desvianza multiplicado por la raíz cuadrada de la inversa del producto del parámetro de escala y 1−influencia del caso.
- **Residuo de verosimilitud.** La raíz cuadrada de un promedio ponderado (basado en la influencia del caso) de los cuadrados de los residuos tipificados de Pearson y de desvianza, con el signo del residuo bruto.

# Modelos lineales generalizados: Exportar

Figura 6-12
*Modelos lineales generalizados: pestaña Exportar*




**Exportar modelo como datos.** Escribe un conjunto de datos de formato IBM® SPSS® Statistics que contiene la matriz de covarianzas o correlaciones de los parámetros con las estimaciones de los parámetros, errores típicos, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **Variables de segmentación.** Si se han utilizado, todas las variables que definan segmentaciones.
- **RowType_.** Toma los valores (y las etiquetas de valor) *COV* (covarianzas), *CORR* (correlaciones), *EST* (estimaciones de los parámetros), *SE* (errores típicos), *SIG* (niveles de significación) y *DF* (grados de libertad del diseño muestral). Hay un caso diferente con el tipo de fila *COV* (o *CORR*) para cada parámetro del modelo, además de un caso diferente para cada uno de los otros tipos de filas.
- **VarName_.** Toma los valores *P1*, *P2*, ..., correspondientes a una lista ordenada de todos los parámetros estimados del modelo (salvo los parámetros binomiales negativos o de escala), para los tipos de fila *COV* o *CORR*, con las etiquetas de valor correspondientes a las cadenas

de parámetros mostradas en la tabla de estimaciones de los parámetros. Las casillas están vacías para los demás tipos de filas.

- **P1, P2, ...** Estas variables corresponden a una lista ordenada de todos los parámetros del modelo (incluidos los parámetros binomiales negativos y de escala, según sea apropiado), con las etiquetas de variable correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros y toman valores según el tipo de fila.

  Para los parámetros redundantes, todas las covarianzas se establecen en cero, las correlaciones se establecen en el valor perdido del sistema; todas las estimaciones de los parámetros se establecen en cero; y todos los errores típicos, niveles de significación y los grados de libertad residuales se establecen en el valor perdido del sistema.

  Para el parámetro de escala, las covarianzas, correlaciones, nivel de significación y grados de libertad se establecen en el valor perdido del sistema. Si el parámetro de escala se estima mediante máxima verosimilitud, se indica el error típico; en otro caso se establece en el valor perdido del sistema.

  Para el parámetro binomial negativo, las covarianzas, correlaciones, nivel de significación y grados de libertad se establecen en el valor perdido del sistema. Si el parámetro binomial negativo se estima mediante máxima verosimilitud, se indica el error típico; en otro caso se establece en el valor perdido del sistema.

  Si hay segmentaciones, se debe acumular la lista de parámetros a través de todas las segmentaciones. En una determinada segmentación, es posible que algunos parámetros sean irrelevantes; pero no es lo mismo que sean redundantes. Para los parámetros irrelevantes, todas las covarianzas y correlaciones, estimaciones de los parámetros, errores típicos, niveles de significación y grados de libertad se establecen en el valor perdido del sistema.

Puede utilizar este archivo matricial como valores iniciales para una estimación posterior del modelo; tenga en cuenta que este archivo no se puede utilizar directamente para realizar otros análisis en otros procedimientos que lean un archivo matricial a menos que dichos procedimientos acepten todos los tipos de filas que aquí se exportan. Incluso en esos casos, deberá asegurarse de que todos los parámetros del archivo matricial tienen el mismo significado para el procedimiento que lee el archivo.

**Exportar modelo como XML.** Guarda las estimaciones de los parámetros y la matriz de covarianzas de los parámetros (si se selecciona) 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 GENLIN*

La sintaxis de comandos también le permite:

- Especificar valores iniciales para las estimaciones de los parámetros como una lista de números (utilizando el subcomando `CRITERIA`).
- Fijar covariables en valores distintos los de sus medias al calcular las medias marginales estimadas (utilizando el subcomando `EMMEANS`).

- Especificar contrastes polinómicos personalizados para las medias marginales estimadas (utilizando el subcomando `EMMEANS`).
- Especificar un subconjunto de los factores para los que se muestran las medias marginales estimadas para compararlos utilizando el tipo de contraste especificado (utilizando las palabras clave `TABLES` y `COMPARE` del subcomando `EMMEANS`).

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

Capítulo
7

# Ecuaciones de estimación generalizadas

El procedimiento Ecuaciones de estimación generalizadas amplía el modelo lineal generalizado para permitir el análisis de medidas repetidas y otras observaciones correlacionadas, como datos conglomerados.

**Ejemplo.** Los funcionarios de la sanidad pública puede utilizar las ecuaciones de estimación generalizadas para ajustar una regresión logística de medidas repetidas para estudiar los efectos de la contaminación del aire sobre los niños.

**Datos.** La respuesta puede ser de escala, de recuentos, binaria o eventos en ensayos. Se supone que los factores son categóricos. Las covariables, el peso de escala y el desplazamiento se suponen que son de escala. Las variables utilizadas para definir los sujetos o las medidas repetidas intra-sujetos no se pueden utilizar para definir la respuesta pero pueden desempeñar otros papeles en el modelo.

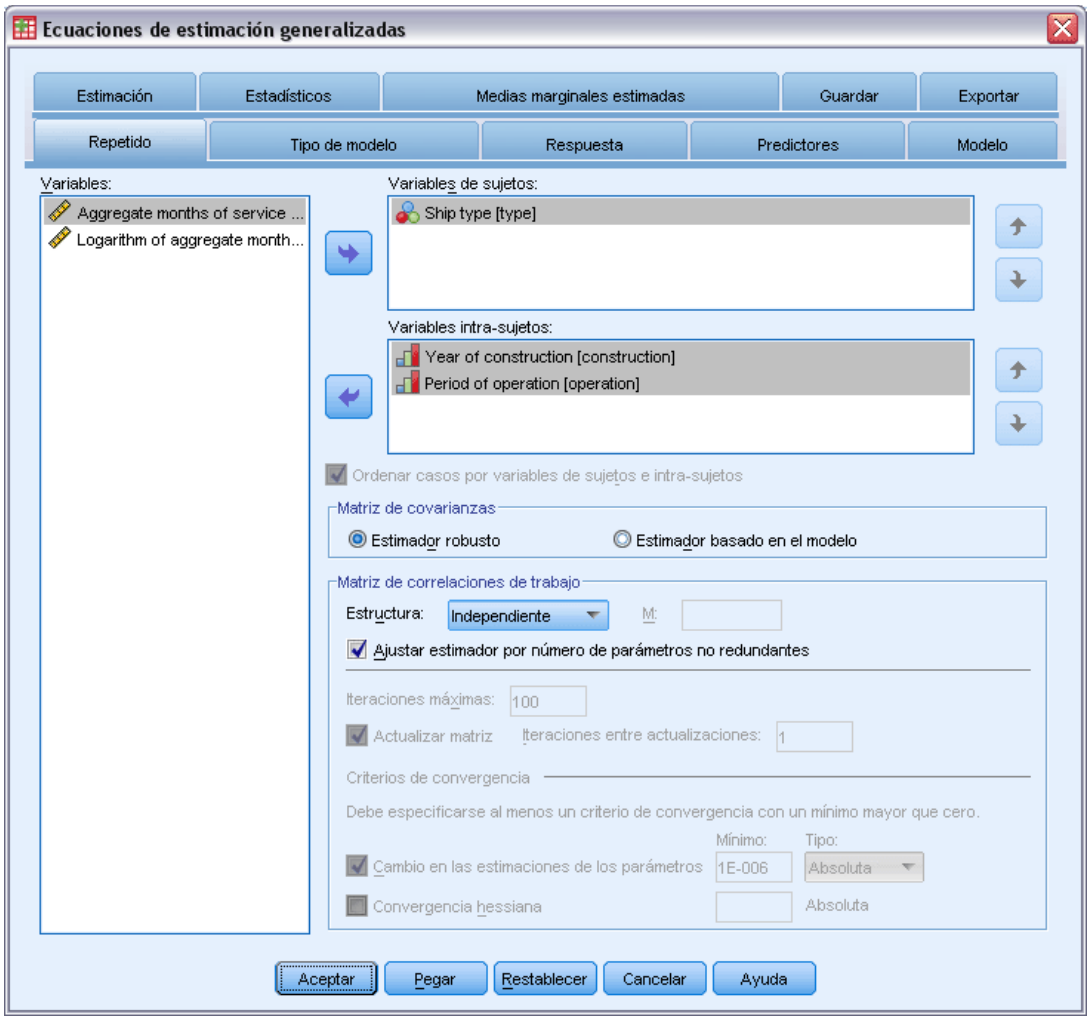
**Supuestos.** Los casos se supone que son dependientes dentro de los sujetos e independientes entre los sujetos. La matriz de correlaciones que representa las dependencias intra-sujetos se estima como parte del modelo.

### *Obtención de ecuaciones de estimación generalizadas*

Seleccione en los menús:

Analizar > Modelos lineales generalizados > Ecuaciones de estimación generalizadas...

Figura 7-1
*Ecuaciones de estimación generalizadas: Pestaña Repetido*




- Seleccione una o más variables de sujeto (más adelante se indican otras opciones).

  La combinación de valores de las variables especificadas debe definir de manera única los **sujetos** del conjunto de datos. Por ejemplo, una única variable *ID de paciente* debería ser suficiente para definir los sujetos de un único hospital, pero puede que sea necesario combinar *ID de hospital* e *ID de paciente* si los números de identificación de paciente no son únicos entre varios hospitales. En una configuración de medidas repetidas, se registran varias observaciones para cada sujeto, de manera que cada sujeto puede ocupar varios casos del conjunto de datos.

- En la pestaña Tipo de modelo, especifique una distribución y una función de enlace.

- En la pestaña Respuesta, seleccione una variable dependiente.

- En la pestaña Predictores, seleccione los factores y las covariables que utilizará para pronosticar la variable dependiente.

▶ En la pestaña Modelo, especifique los efectos del modelo utilizando las covariables y los factores seleccionados.

Si lo desea, en la pestaña Repetido puede especificar:

**Variables intra-sujetos.** La combinación de valores de las variables intra-sujetos define el orden de las medidas dentro de los sujetos. Por tanto, la combinación de las variables intra-sujetos y de los sujetos define de manera única cada medida. Por ejemplo, la combinación de *Período*, *ID de hospital* e *ID de paciente* define, para cada caso, una determinada visita a la consulta de un determinado paciente dentro de un determinado hospital.

Si el conjunto de datos ya está ordenado de manera que las medidas repetidas de cada sujeto se producen en un bloque contiguo de casos y en el orden correcto, no es estrictamente necesario especificar un variable intra-sujetos y puede anular la selección de Ordenar casos por variables de sujetos e intra-sujetos con el fin de ahorrar el tiempo de procesamiento necesario para determinar el orden (temporal). Por lo general, es aconsejable utilizar las variables intra-sujetos para asegurarse de que las medidas se ordenan correctamente.

Las variables de sujetos e intra-sujetos no se pueden utilizar para definir la respuesta, pero pueden realizar otras funciones en el modelo. Por ejemplo, *ID de hospital* se puede utilizar como factor en el modelo.

**Matriz de covarianzas.** El estimador basado en el modelo es la negativa de la inversa generalizada de la matriz hessiana. El estimador robusto (también llamado el estimador de Huber/White/Sandwich) es un estimador basado en el modelo "corregido" que proporciona una estimación coherente de la covarianza, incluso cuando la matriz de correlaciones de trabajo se especifica incorrectamente. Esta especificación se aplica a los parámetros del modelo lineal que forma parte de las ecuaciones de estimación generalizadas, mientras que la especificación de la pestaña Estimación se aplica únicamente al modelo lineal generalizado inicial.

**Matriz de correlaciones de trabajo.** Esta matriz de correlaciones representa las dependencias intra-sujetos. Su tamaño queda determinado por el número de medidas y, por tanto, por la combinación de los valores de las variables intra-sujetos. Puede especificar una de las siguientes estructuras:

- **Independiente.** Las medidas repetidas no están correlacionadas.
- **AR(1).** Las medidas repetidas tienen una relación autorregresiva de primer orden. La correlación entre dos elementos es igual a $\rho$ en el caso de elementos adyacentes, $\rho^2$ cuando se trata de elementos separados entre sí por un tercero, y así sucesivamente. $\rho$ está limitado de manera que $-1<\rho<1$.
- **Intercambiable.** Esta estructura tiene correlaciones homogéneas entre los elementos. También se conoce como una estructura de simetría compuesta.
- **M-dependiente.** Las medidas consecutivas tienen un coeficiente de correlación común, pares de medidas separadas por una tercera tienen un coeficiente de correlación común y así sucesivamente, hasta pares de medidas separadas por $m-1$ otras medidas. Se supone que las medidas con una mayor separación no están correlacionadas. Al elegir esta estructura, especifique un valor de $m$ que sea menor que el orden de la matriz de correlaciones de trabajo.
- **Sin estructura.** Es una matriz de correlaciones completamente general.

Por defecto, el procedimiento ajustará las estimaciones de correlación utilizando el número de parámetros que no sean redundantes. Puede que sea aconsejable eliminar este ajuste si desea que las estimaciones sean invariables frente a los cambios de replicación a nivel de sujeto en los datos.

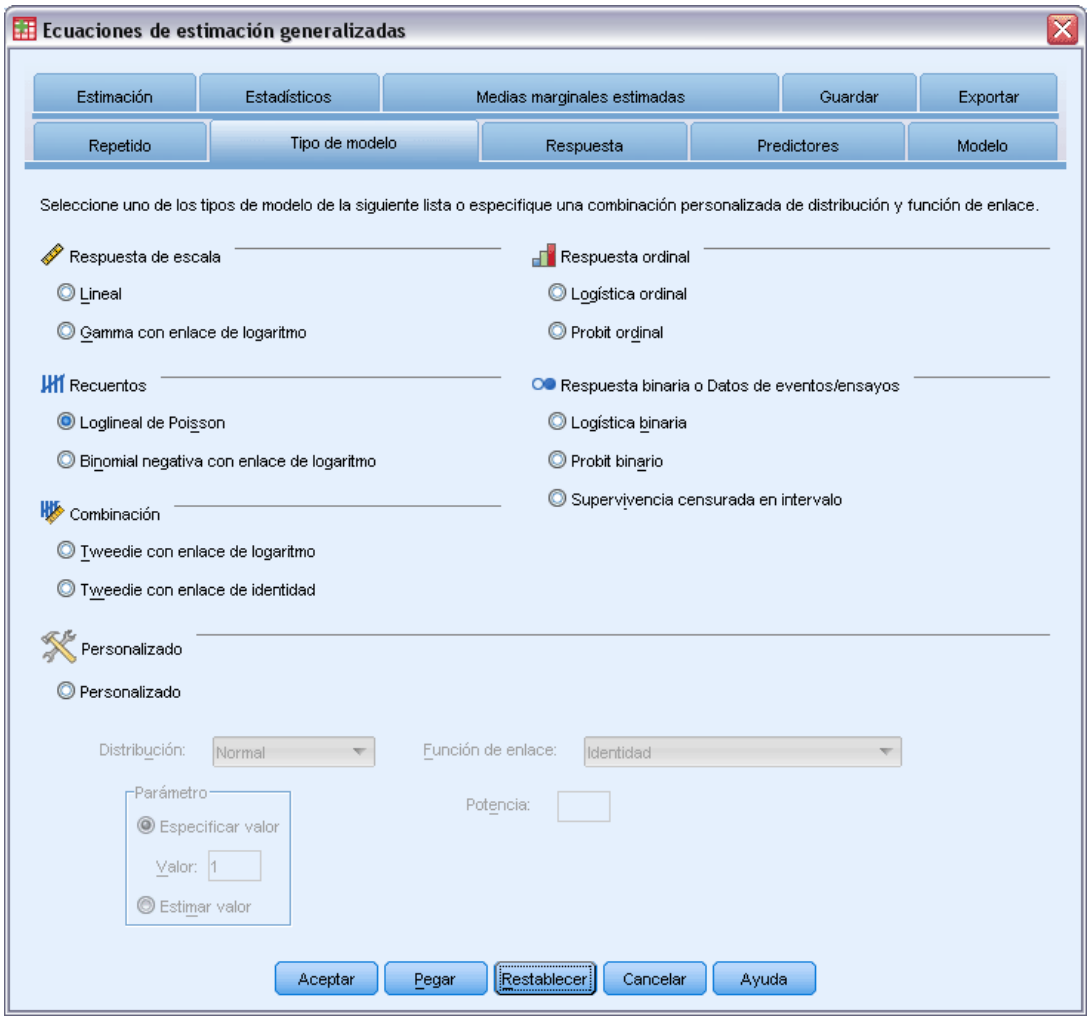
- **Número máximo de iteraciones.** Número máximo de iteraciones que ejecutará el algoritmo de ecuaciones de estimación generalizadas. Especifique un número entero no negativo. Esta especificación se aplica a los parámetros del modelo lineal que forma parte de las ecuaciones de estimación generalizadas, mientras que la especificación de la pestaña Estimación se aplica únicamente al modelo lineal generalizado inicial.
- **Actualizar matriz.** Los elementos de la matriz de correlaciones de trabajo se estiman basándose en las estimaciones de los parámetros, que se actualizan en cada iteración del algoritmo. Si la matriz de correlaciones de trabajo no se actualiza en absoluto, se utilizará la matriz de correlaciones de trabajo inicial en todo el proceso de estimación. Si se actualiza la matriz, puede especificar el intervalo de iteración según el que se actualizarán los elementos de la matriz de correlaciones de trabajo. La especificación de un valor mayor que 1 puede reducir el tiempo de procesamiento.

**Criterios de convergencia.** Estas especificaciones se aplican a los parámetros del modelo lineal que forma parte de las ecuaciones de estimación generalizadas, mientras que la especificación de la pestaña Estimación se aplica únicamente al modelo lineal generalizado inicial.

- **Convergencia de los parámetros.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en las estimaciones de los parámetros son inferiores al valor especificado, que debe ser positivo.
- **Convergencia hessiana.** Se asume la convergencia si un estadístico basado en la hessiana es inferior al valor especificado, que debe ser positivo.

# Ecuaciones de estimación generalizadas: Tipo de modelo

Figura 7-2
*Ecuaciones de estimación generalizadas: Pestaña Tipo de modelo*




La pestaña Tipo de modelo permite especificar la distribución y la función de enlace del modelo, además de proporcionar accesos directos a varios modelos habituales que aparecen clasificados por tipo de respuesta.

### *Tipos de modelos*

**Respuesta de escala.**

- **Lineal.** Especifica la distribución normal y la función de enlace identidad.
- **Gamma con enlace de logaritmo.** Especifica la distribución gamma y la función de enlace de logaritmo.

**Respuesta ordinal.**

- **Logística ordinal.** Especifica la distribución multinomial (ordinal) y la función de enlace logit acumulado.
- **Probit ordinal.** Especifica la distribución multinomial (ordinal) y la función de enlace probit acumulado.

**Recuentos.**

- **Loglineal de Poisson.** Especifica la distribución de Poisson y la función de enlace de logaritmo.
- **Binomial negativa con enlace de logaritmo.** Especifica la distribución binomial negativa (con el valor 1 para el parámetro auxiliar) y la función de enlace de logaritmo. Para que el procedimiento calcule el valor del parámetro auxiliar, especifique un modelo personalizado con distribución binomial negativa y seleccione Estimar valor en el grupo de parámetros.

**Respuesta binaria o Datos de eventos/ensayos.**

- **Logística binaria.** Especifica la distribución binomial y la función de enlace logit.
- **Probit binario.** Especifica la distribución binomial y la función de enlace probit.
- **Supervivencia censurada en intervalo.** Especifica la distribución binomial y la función de enlace log-log complementario.

**Combinación.**

- **Tweedie con enlace de logaritmo.** Especifica la distribución de Tweedie y la función de enlace de logaritmo.
- **Tweedie con enlace de identidad.** Especifica la distribución de Tweedie y la función de enlace identidad.

**Personalizado.** Especifique su propia combinación de distribución y función de enlace.

***Distribución***

Esta selección especifica la distribución de la variable dependiente. La posibilidad de especificar una distribución que no sea la normal y una función de enlace que no sea la identidad es la principal mejora que aporta el modelo lineal generalizado respecto al modelo lineal general. Hay muchas combinaciones posibles de distribución y función de enlace, varias de las cuales pueden ser adecuadas para un determinado conjunto de datos, por lo que su elección puede estar guiada por consideraciones teóricas a priori y por las combinaciones que parezcan funcionar mejor.

- **Binomial.** Esta distribución es adecuada únicamente para las variables que representan una respuesta binaria o un número de eventos.
- **Gamma.** Esta distribución es adecuada para las variables con valores de escala positivos que se desvían hacia valores positivos más grandes. Si un valor de datos es menor o igual que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **De Gauss inversa.** Esta distribución es adecuada para las variables con valores de escala positivos que se desvían hacia valores positivos más grandes. Si un valor de datos es menor o igual que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **Binomial negativa.** Esta distribución considera el número de intentos necesarios para lograr *k* éxitos y es adecuada para variables que tengan valores enteros que no sean negativos. Si un valor de datos no es entero, es menor que 0 o es un valor perdido, el correspondiente caso no

se utilizará en el análisis. El valor del parámetro auxiliar de la distribución binomial negativa puede ser cualquier número mayor o igual que 0; se puede establecer en un valor fijo o dejar que lo estime el procedimiento. Cuando el parámetro auxiliar se establece en 0, utilizar esta distribución equivale a utilizar la distribución de Poisson.

- **Normal.** Es adecuada para variables de escala cuyos valores adoptan una distribución simétrica con forma de campana en torno a un valor central (la media). La variable dependiente debe ser numérica.
- **Poisson.** Esta distribución considera el número de ocurrencias de un evento de interés en un período fijo de tiempo y es apropiada para variables que tengan valores enteros que no sean negativos. Si un valor de datos no es entero, es menor que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **Tweedie.** Esta distribución es adecuada para variables que puedan representarse mediante mezclas de Poisson de distribuciones gamma; la distribución es una "mezcla" en el sentido de que combina las propiedades de distribuciones continuas (toma valores reales no negativos) y discretas (masa de probabilidad positiva en un único valor, 0). La variable dependiente debe ser numérica y los valores de los datos deben ser iguales o mayores que cero. Si un valor de datos es menor que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis. El valor fijo del parámetro de la distribución de Tweedie puede ser cualquier número mayor que uno y menor que dos.
- **Multinomial.** Esta distribución es adecuada para variables que representan una respuesta ordinal. La variable dependiente puede ser numérica o de cadena, y debe tener como mínimo dos valores válidos distintos de los datos.

### *Función de enlace*

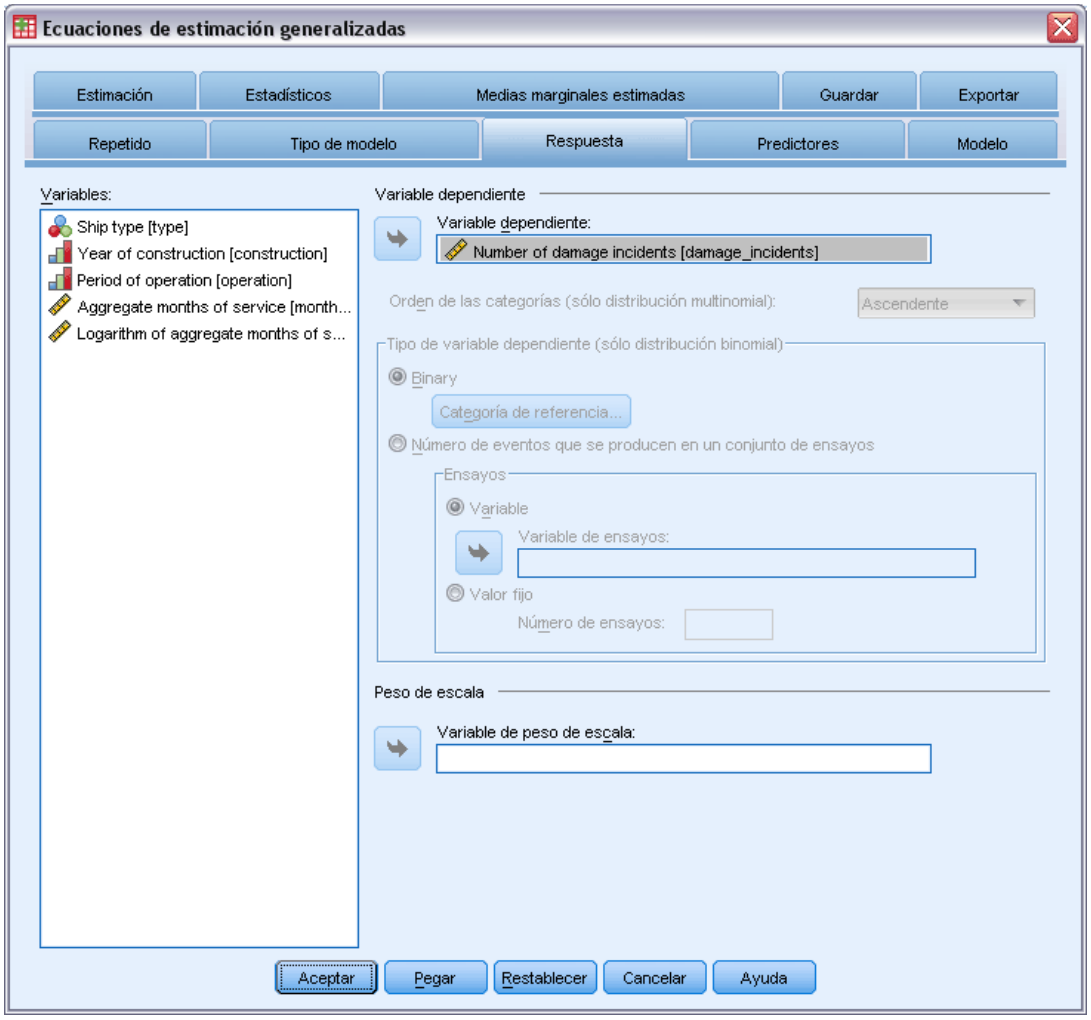
La función de enlace es una transformación de la variable dependiente que permite la estimación del modelo. Se encuentran disponibles las siguientes funciones:

- **Identidad.** $f(x)=x$. No se transforma la variable dependiente. Este vínculo se puede utilizar con cualquier distribución.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Es apropiada únicamente para la distribución binomial.
- **Cauchit acumulada.** $f(x) = \tan(\pi\ (x - 0.5))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Log-log complementario acumulado.** $f(x)=\ln(-\ln(1-x))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Logit acumulado.** $f(x)=\ln(x / (1-x))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta.. Es apropiada únicamente para la distribución multinomial.
- **Log-log negativo acumulado.** $f(x)=(-\ln(-\ln x))$, aplicada a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Probit acumulada.** $f(x)=\Phi^{-1}(x)$, aplicada a la probabilidad acumulada de cada categoría de la respuesta, donde $\Phi^{-1}$ es la función de distribución acumulada normal típica inversa. Es apropiada únicamente para la distribución multinomial.
- **Log.** $f(x)=\log(x)$. Este vínculo se puede utilizar con cualquier distribución.
- **Complemento log.** $f(x)=\log(1-x)$. Es apropiada únicamente para la distribución binomial.

- **Logit.** $f(x)=\log(x / (1-x))$. Es apropiada únicamente para la distribución binomial.
- **Binomial negativa.** $f(x)=\log(x / (x+k^{-1}))$, donde $k$ es el parámetro auxiliar de la distribución binomial negativa. Es apropiada únicamente para la distribución binomial negativa.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Es apropiada únicamente para la distribución binomial.
- **Potencia de las ventajas.** $f(x)=[(x/(1-x))^{\alpha}-1]/\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. $\alpha$ es la especificación de número necesaria y debe ser un número real. Es apropiada únicamente para la distribución binomial.
- **Probit.** $f(x)=\Phi^{-1}(x)$, donde $\Phi^{-1}$ es la función de distribución acumulada normal típica inversa. Es apropiada únicamente para la distribución binomial.
- **Potencia.** $f(x)=x^{\alpha}$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. $\alpha$ es la especificación de número necesaria y debe ser un número real. Este vínculo se puede utilizar con cualquier distribución.

## Respuesta de las ecuaciones de estimación generalizadas

Figura 7-3
*Ecuaciones de estimación generalizadas: Pestaña Respuesta*



En muchos casos, puede especificar sencillamente una variable dependiente. No obstante, las variables que adoptan únicamente dos valores y las respuestas que registran eventos en ensayos que requieren una atención adicional.

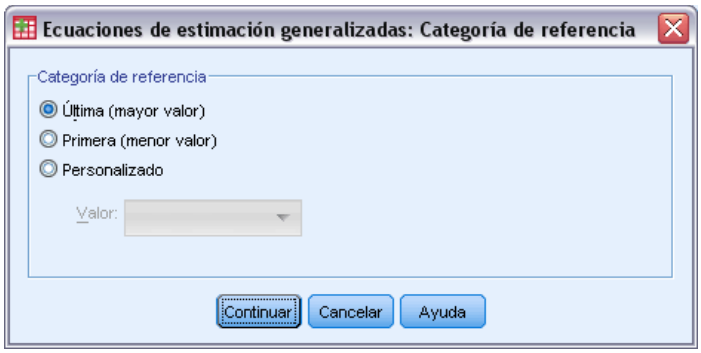
- **Respuesta binaria.** Cuando la variable dependiente adopta únicamente dos valores, puede especificar la categoría de referencia para la estimación de los parámetros. Una variable de respuesta binaria puede ser de cadena o numérica.
- **Número de eventos que se producen en un conjunto de ensayos** Cuando la respuesta es un número de eventos que ocurren en un conjunto de ensayos, la variable dependiente contiene el número de eventos y puede seleccionar una variable adicional que contenga el número de ensayos. Otra posibilidad, si el número de ensayos es el mismo en todos los sujetos, consiste en especificar los ensayos mediante un valor fijo. El número de ensayos debe ser mayor o igual que el número de eventos para cada caso. Los eventos deben ser enteros no negativos y los ensayos deben ser enteros positivos.

Para los modelos multinomiales ordinales, puede especificar el orden de las categorías de la respuesta: ascendente, descendente o datos (el orden de los datos indica que el primer valor encontrado en los datos define la primera categoría y el último valor encontrado define la última categoría).

**Peso de escala.** El parámetro de escala es un parámetro del modelo estimado relacionado con la varianza de la respuesta. Los pesos de escala son valores "conocidos" que pueden variar de una observación a otra. Si se especifica una variable de peso de escala, el parámetro de escala, que está relacionado con la varianza de la respuesta, se divide por él para cada observación. Los casos cuyo valor del peso de escala es menor o igual que 0 o que son perdidos no se utilizan en el análisis.

## *Ecuaciones de estimación generalizadas: Categoría de referencia*

Figura 7-4
*Cuadro de diálogo Ecuaciones de estimación generalizadas: Categoría de referencia*

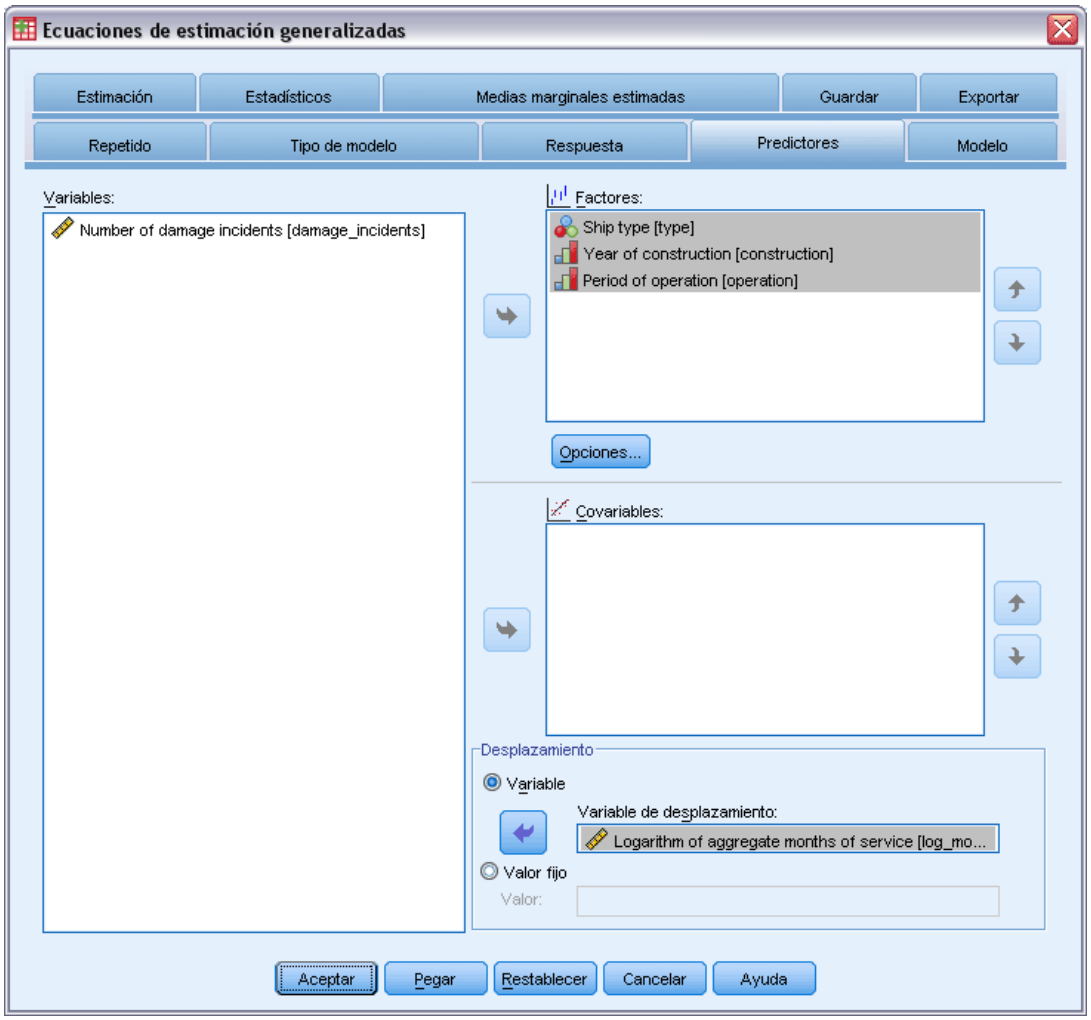


Para una respuesta binaria, puede elegir la categoría de referencia de la variable dependiente. Puede afectar a ciertos resultados, como las estimaciones de los parámetros y los valores guardados, pero no debería cambiar el ajuste del modelo. Por ejemplo, si la respuesta binaria toma los valores 0 y 1:

- Por defecto, el procedimiento utiliza la última categoría (la de mayor valor), o 1, como la categoría de referencia. En esta situación, las probabilidades guardadas por el modelo estiman la posibilidad de que un determinado caso tome el valor 0 y las estimaciones de los parámetros deben interpretarse como relativas a la probabilidad de la categoría 0.
- Si especifica la primera categoría (la de menor valor), o 0, como la categoría de referencia, las probabilidades guardadas por el modelo estimarán la posibilidad de que un determinado caso tome el valor 1.
- Si especifica la categoría personalizada y la variable tiene etiquetas definidas, puede establecer la categoría de referencia eligiendo un valor de la lista. Puede resultar cómodo si, mientras se especifica un modelo, no recuerda exactamente cómo se ha codificado una determinada variable.

# Ecuaciones de estimación generalizadas: Predictores

Figura 7-5
*Ecuaciones de estimación generalizadas: Pestaña Predictores*




La pestaña Predictores permite especificar los factores y las covariables que se utilizarán para crear los efectos del modelo y para especificar un desplazamiento opcional.

**Factores.** Los factores son predictores categóricos y pueden ser numéricos o de cadena.

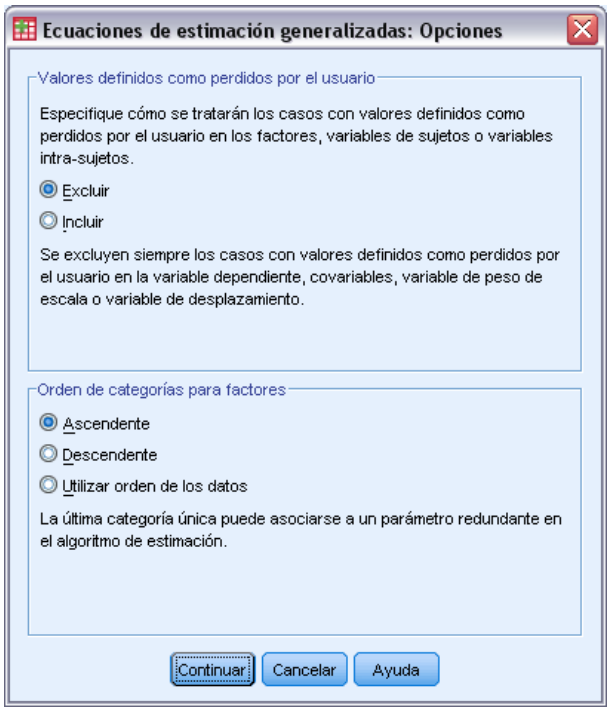
**Covariables.** Las covariables son predictores de escala y deben ser numéricas.

*Nota*: Cuando la respuesta es binomial con formato binario, el procedimiento calcula los estadísticos de bondad de ajuste de chi cuadrado y de desvianza por subpoblaciones que se basan en la clasificación cruzada de los valores observados de los factores y las covariables seleccionadas. Debe mantener el mismo conjunto de predictores en las diferentes ejecuciones del procedimiento para asegurarse de que se utiliza un número coherente de subpoblaciones.

**Desplazamiento.** El término desplazamiento es un predictor "estructural". Su coeficiente no se estima por el modelo pero se supone que tiene el valor 1. Por tanto, los valores del desplazamiento se suman sencillamente al predictor lineal de la variable dependiente. Esto resulta especialmente útil en los modelos de regresión de Poisson, en los que cada caso pueden tener diferentes niveles de exposición al evento de interés. Por ejemplo, al modelar las tasas de accidente de diferentes conductores, hay una importante diferencia entre un conductor que ha sido el culpable de un accidente en tres años y un conductor que ha sido el culpable de un accidente en 25 años. El número de accidentes se puede modelar como una respuesta de Poisson si la experiencia del conductor se incluye como un término de desplazamiento.

## *Ecuaciones de estimación generalizadas: Opciones*

Figura 7-6
*Cuadro de diálogo Ecuaciones de estimación generalizadas: Opciones*




Estas opciones se aplican a todos los factores especificados en la pestaña Predictores.

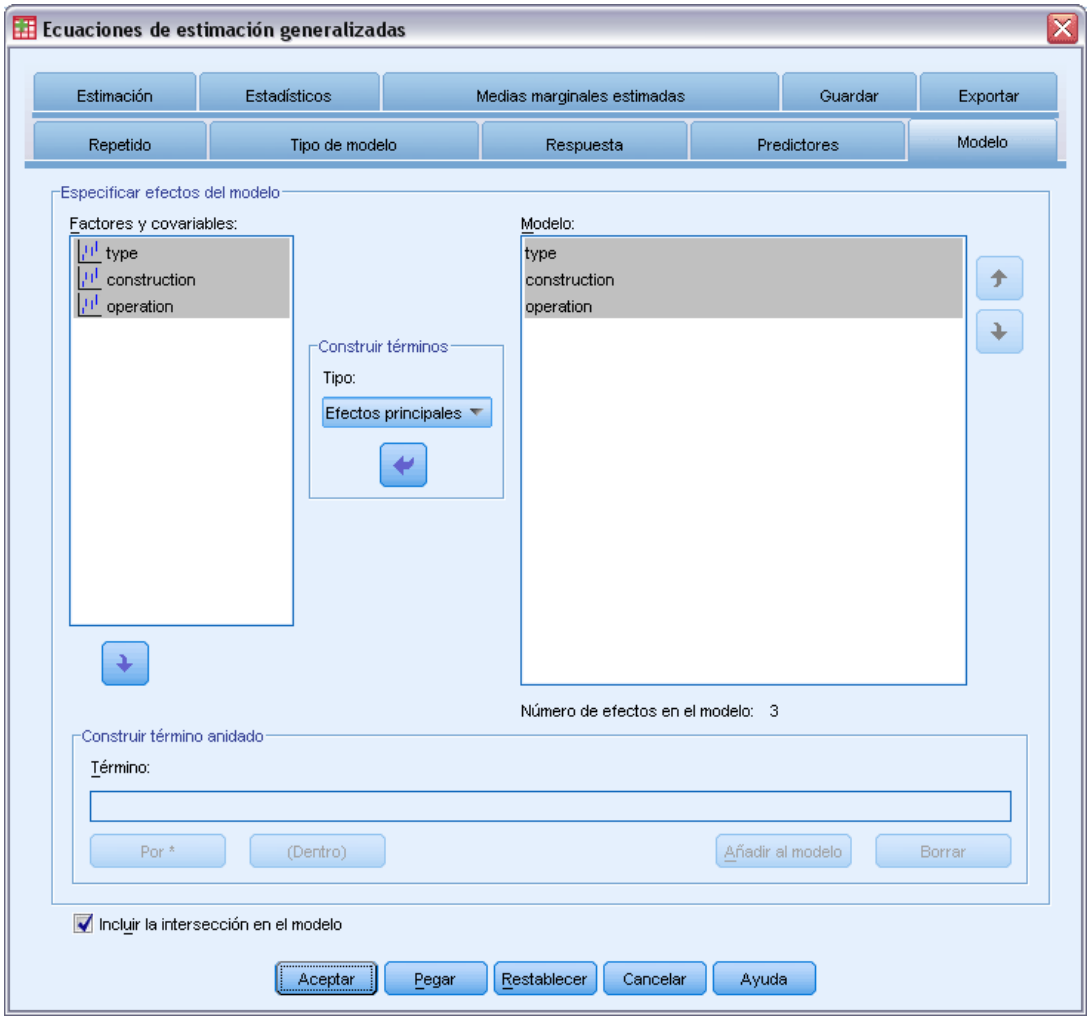
**Valores definidos como perdidos por el usuario.** Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores definidos como perdidos por el usuario se deben tratar como válidos entre las variables de factor.

**Orden de categorías.** Es relevante para determinar el último nivel de un factor, que puede estar asociado a un parámetro redundante del algoritmo de estimación. Si se cambia el orden de categorías es posible que cambien también los valores de los efectos de los niveles de los factores, ya que estas estimaciones de los parámetros se calculan respecto al "último" nivel. Los factores se pueden ordenar en orden ascendente desde el valor mínimo hasta el máximo, en orden descendente desde el valor máximo hasta el mínimo o siguiendo el "orden de los datos".

Significa que el primer valor encontrado en los datos define la primera categoría, y el último valor único encontrado define la última categoría.

# Ecuaciones de estimación generalizadas: Modelo

Figura 7-7
*Ecuaciones de estimación generalizadas: Pestaña Modelo*




**Especificar efectos del modelo.** El modelo por defecto sólo utiliza la intersección, por lo que deberá especificar explícitamente todos los demás efectos del modelo. Puede elegir entre términos anidados o no anidados.

### *Términos no anidados*

Para las covariables y los factores seleccionados:

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

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

**Factorial.** Crea todas las interacciones y efectos principales posibles para las variables seleccionadas.

**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.

### *Términos anidados*

En este procedimiento, puede construir términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción o añadir varios niveles de anidación al término anidado.

**Limitaciones.** Existen las siguientes restricciones para los términos anidados:

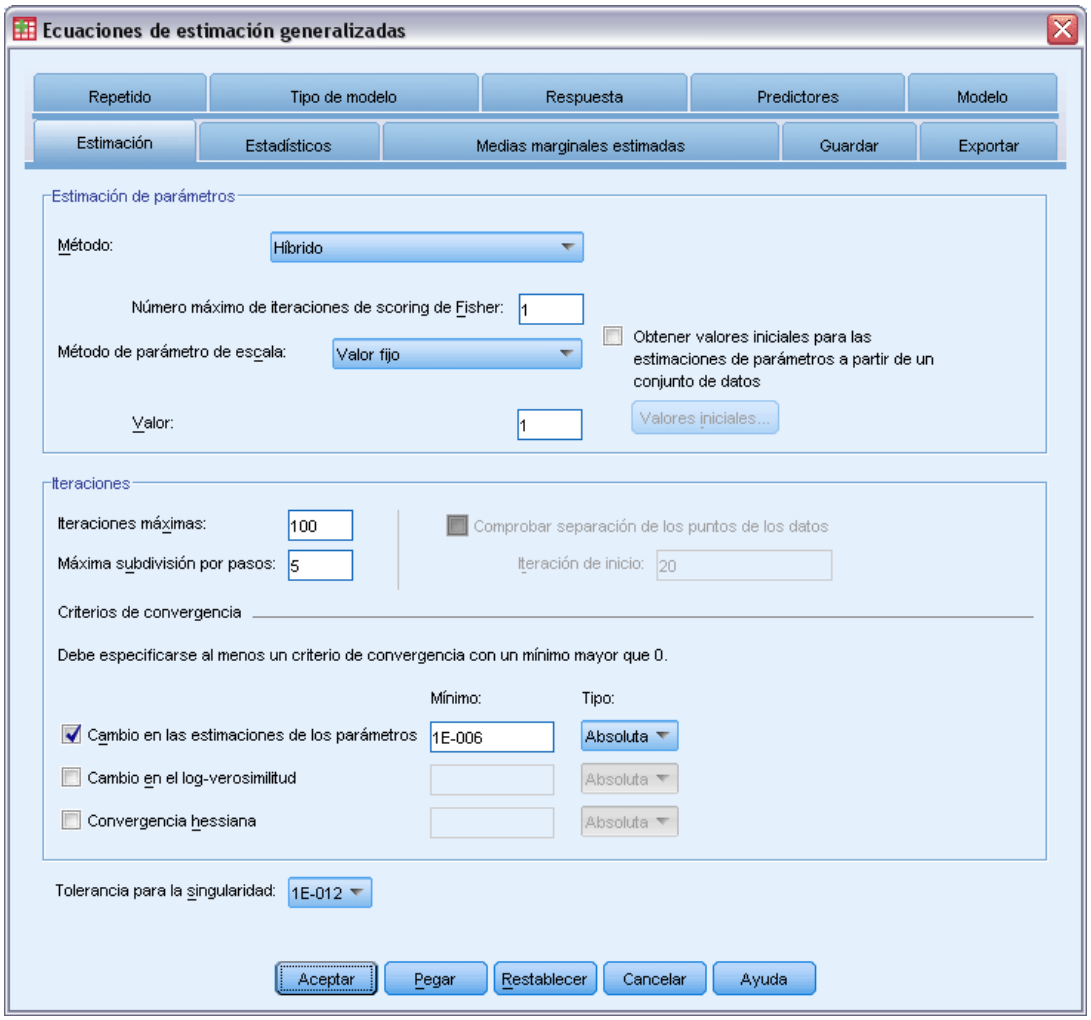
- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A*A*.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A(A)*.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si *A* es un factor y *X* es una covariable, no es válido especificar *A(X)*.

**Intersección.** La intersección se incluye normalmente en el modelo. Si asume que los datos pasan por el origen, puede excluir la intersección.

Los modelos con distribución ordinal multinomial no tienen un único término de intersección, sino que tienen parámetros de umbral que definen los puntos de transición entre las categorías adyacentes. Los umbrales siempre se incluyen en el modelo.

# Ecuaciones de estimación generalizadas: Estimación

Figura 7-8
*Ecuaciones de estimación generalizadas: Pestaña Estimación*




**Estimación de parámetros.** Los controles de este grupo le permiten especificar los métodos de estimación y proporcionar los valores iniciales para las estimaciones de los parámetros.

- **Método.** Puede seleccionar un método de estimación de parámetros. Los métodos disponibles son Newton-Raphson, Scoring de Fisher o un método híbrido en el que las iteraciones de Scoring de Fisher se realizan antes de cambiar al método de Newton-Raphson. Si se logra la convergencia durante la fase de Scoring de Fisher del método híbrido antes de que se lleven a cabo el número máximo de iteraciones de Fisher, el algoritmo continúa con el método de Newton-Raphson.
- **Método de parámetro de escala.** Puede seleccionar el método de estimación del parámetro de escala.

La máxima verosimilitud estima conjuntamente el parámetro de escala y los efectos del modelo. Tenga en cuenta que esta opción no es válida si la respuesta tiene una distribución binomial negativa, de Poisson o binomial. Como el concepto de verosimilitud no encaja en las ecuaciones de estimación generalizadas, esta especificación se aplica únicamente al modelo lineal generalizado inicial. A continuación, esta estimación del parámetro de escala se pasa a las ecuaciones de estimación generalizadas, que actualizan el parámetro de escala con el chi-cuadrado de Pearson dividido por sus grados de libertad.

Las opciones de desvianza y de chi-cuadrado de Pearson estiman el parámetro de escala a partir del valor de dichos estadísticos del modelo lineal generalizado inicial. A continuación, esta estimación del parámetro de escala se pasa a las ecuaciones de estimación generalizadas, que lo tratan como corregido.

Otra posibilidad consiste en especificar un valor corregido para el parámetro de escala. Se tratará como corregido al estimar el modelo lineal generalizado inicial y las ecuaciones de estimación generalizadas.

- **Valores iniciales.** El procedimiento calculará automáticamente los valores iniciales de los parámetros. Otra posibilidad consiste en especificar los valores iniciales de las estimaciones de los parámetros.

Las iteraciones y los criterios de convergencia especificados en esta pestaña se aplican únicamente al modelo lineal generalizado inicial. Para ver los criterios de estimación utilizados para ajustar las ecuaciones de estimación generalizadas, consulte la pestaña Repetida.

**Iteraciones.**

- **Iteraciones máximas.** Número máximo de iteraciones que se ejecutará el algoritmo. Especifique un número entero no negativo.
- **Máxima subdivisión por pasos.** En cada iteración, se reduce el tamaño del paso mediante un factor de 0,5 hasta que aumenta el logaritmo de la verosimilitud o se alcanza la máxima subdivisión por pasos. Especifique un número entero positivo.
- **Comprobar si hay separación completa de los puntos de los datos.** Si se activa, el algoritmo realiza una prueba para garantizar que las estimaciones de los parámetros tienen valores exclusivos. Se produce una separación cuando el procedimiento pueda generar un modelo que clasifique cada caso de forma correcta. Esta opción no está disponible para respuestas multinomiales y binomiales con formato binario.

**Criterios de convergencia.**

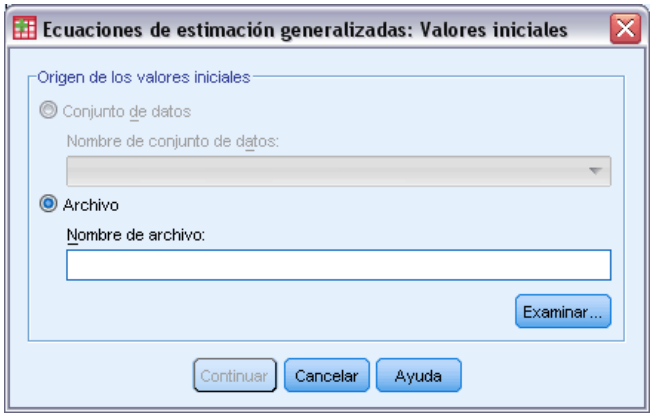
- **Convergencia de los parámetros.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en las estimaciones de los parámetros son inferiores al valor especificado, que debe ser positivo.
- **Convergencia del logaritmo de la verosimilitud.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en la función de log-verosimilitud sean inferiores que el valor especificado, que debe ser positivo.
- **Convergencia hessiana.** En el caso de la especificación Absoluta, se supone la convergencia si un estadístico basado en la convergencia hessiana es menor que el valor positivo especificado. En el caso de la especificación Relativa, se supone la convergencia si el estadístico es menor que el producto del valor positivo especificado y el valor absoluto del logaritmo de la verosimilitud.

**Tolerancia para la singularidad.** Las matrices singulares (que no se pueden invertir) tienen columnas linealmente dependientes, lo que causar graves problemas al algoritmo de estimación. Incluso las matrices casi singulares pueden generar resultados deficientes, por lo que el procedimiento tratará una matriz cuyo determinante es menor que la tolerancia como singular. Especifique un valor positivo.

## Ecuaciones de estimación generalizadas: Valores iniciales

El procedimiento estima un modelo lineal generalizado inicial y las estimaciones de este modelo se utilizan como valores iniciales para las estimaciones de los parámetros en la parte de modelo lineal de las ecuaciones de estimación generalizadas. Los valores iniciales no son necesarios para la matriz de correlaciones de trabajo, ya que los elementos de la matriz se basan en las estimaciones de los parámetros. Los valores iniciales especificados en este cuadro de diálogo se utilizan como punto de partida del modelo lineal generalizado inicial, no las ecuaciones de estimación generalizadas, a menos que el número máximo de iteraciones establecido en la pestaña Estimación esté definido como 0.

Figura 7-9
*Cuadro de diálogo Ecuaciones de estimación generalizadas: Valores iniciales*




Si se especifican valores iniciales, deben proporcionarse para todos los parámetros del modelo (incluidos los parámetros redundantes). En el conjunto de datos, el orden de las variables de izquierda a derecha debe ser: *TipoFila_*, *NombreVar_*, *P1*, *P2*, …, donde *TipoFila_* y *NombreVar_* son variables de cadena y *P1*, *P2*, … son variables numéricas que corresponden a una lista ordenada de los parámetros.

- Los valores iniciales se proporcionan en un registro con el valor *EST* para la variable *TipoFila_*; los valores iniciales reales se proporcionan en las variables *P1*, *P2*, …. El procedimiento ignora todos los registros para los que *TipoFila_* tienen un valor diferente de *EST*, así como todos los registros posteriores a la primera aparición de *TipoFila_* igual a *EST*.
- La intersección, si se incluye en el modelo, o los parámetros de umbral, si la respuesta sigue una distribución multinomial, deben ser los primeros valores iniciales.

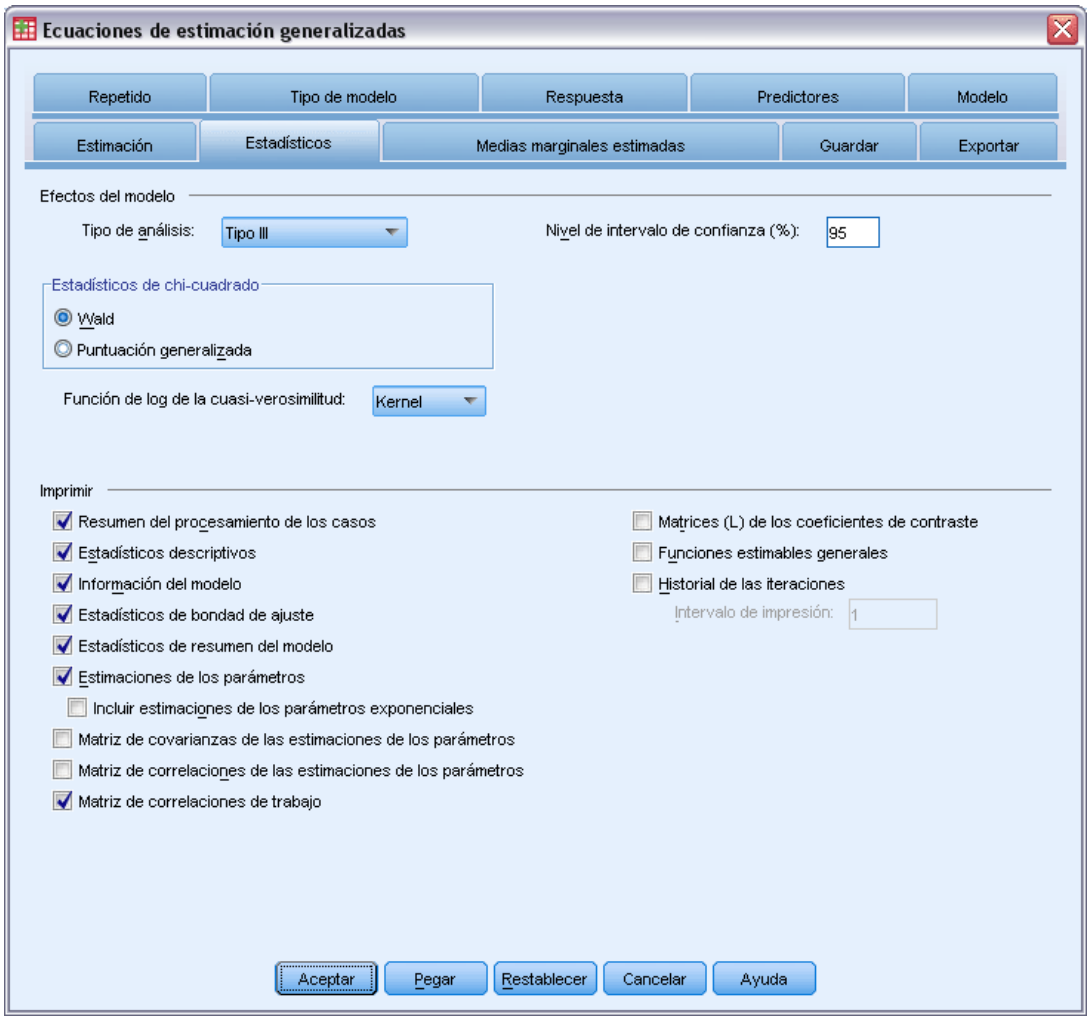
- El parámetro de escala y, si la respuesta sigue una distribución binomial negativa, el parámetro binomial negativo, deben ser los últimos valores iniciales especificados.
- Si está activo Segmentar archivo, las variables deberán comenzar con la variable (o las variables) de segmentación del archivo en el orden especificado al crear la segmentación del archivo, seguidas de *TipoFila_*, *NombreVar_*, *P1*, *P2*, … como se ha indicado anteriormente. La segmentación debe haberse realizado en el conjunto de datos especificado en el mismo orden que en el conjunto de datos original.

*Nota*: Los nombres de las variables *P1*, *P2*, … no son necesarios. El procedimiento aceptará cualquier nombre de variable válido para los parámetros, ya que la asignación de las variables a los parámetros se basa en la posición de la variable y no en el nombre de la variable. Se ignorarán todas las variables que aparezcan después del último parámetro.

La estructura de archivo de los valores iniciales es la misma que la utilizada al exportar el modelo como datos. Por tanto, puede utilizar los valores finales de una ejecución del procedimiento como entrada de una ejecución posterior.

# Ecuaciones de estimación generalizadas: Estadísticos

Figura 7-10
*Ecuaciones de estimación generalizadas: Pestaña Estadísticos*




**Efectos del modelo.**

- **Tipo de análisis** Especifique el tipo de análisis que desea generar para contrastar los efectos del modelo. El análisis de tipo I suele ser apropiado cuando tiene motivos a priori para ordenar los predictores del modelo, mientras que el tipo III es de aplicación más general. Los estadísticos generalizados de puntuación o de Wald se calculan a partir de la selección realizada en el grupo Estadísticos de chi-cuadrado.

- **Intervalos de confianza.** Especifique un nivel de confianza mayor que 50 y menor que 100. Los intervalos de Wald se generan siempre independientemente del tipo de estadísticos de chi-cuadrado seleccionado y se basan el supuesto de que los parámetros siguen una distribución normal asintótica.
- **Función de log de la cuasi-verosimilitud.** Controla el formato de presentación de la función de log de la cuasi-verosimilitud. La función completa incluye un término adicional que es constante respecto a las estimaciones de los parámetros. No tiene ningún efecto sobre la estimación de los parámetros y se deja fuera de la presentación en algunos productos de software.

**Imprimir.** Los siguientes resultados están disponibles.

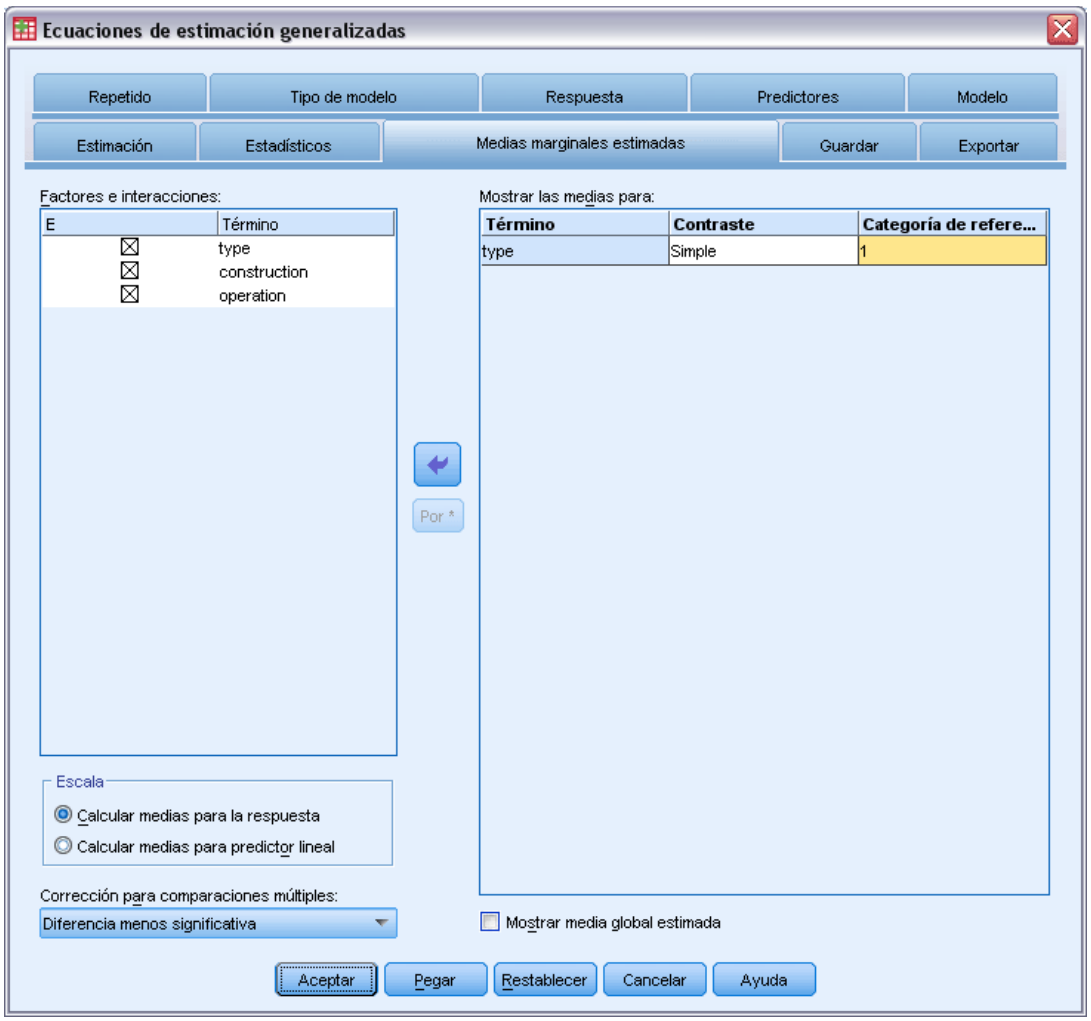
- **Resumen de procesamiento de casos.** Muestra el número y el porcentaje de los casos incluidos y excluidos del análisis y la tabla Resumen de datos correlacionados.
- **Estadísticos descriptivos.** Muestra estadísticos descriptivos e información resumida acerca de los factores, las covariables y la variable dependiente.
- **Información del modelo.** Muestra el nombre del conjunto de datos, la variable dependiente o las variables de eventos y ensayos, la variable de desplazamiento, la distribución de probabilidad y la función de enlace.
- **Estadísticos de bondad de ajuste.** Muestra dos extensiones del criterio de información de Akaike para la selección del modelo: Criterio de cuasi-verosimilitud bajo el modelo de independencia (QIC) para elegir la mejor estructura de correlación y otra medida de QIC para elegir el mejor subconjunto de predictores.
- **Estadísticos de resumen del modelo.** Muestra contraste de ajuste del modelo, incluidos los estadísticos de la razón de la verosimilitud para el contraste ómnibus del ajuste del modelo y los estadísticos para los contrastes de tipo I o III para cada efecto.
- **Estimaciones de los parámetros.** Muestra las estimaciones de los parámetros y los correspondientes estadísticos de contraste e intervalos de confianza. Si lo desea, puede mostrar las estimaciones exponenciadas de los parámetros además de las estimaciones brutas de los parámetros.
- **Matriz de covarianzas de las estimaciones de los parámetros.** Muestra la matriz de covarianzas de los parámetros estimados.
- **Matriz de correlaciones de las estimaciones de los parámetros.** Muestra la matriz de correlaciones de los parámetros estimados.
- **Matrices (L) de los coeficientes de contraste.** Muestra los coeficientes de los contrastes para los efectos por defecto y para las medias marginales estimadas, si se solicitaron en la pestaña Medias marginales estimadas.
- **Funciones estimables generales.** Muestra las matrices para generar las matrices (L) de los coeficientes de contraste.
- **Historial de iteraciones.** Muestra el historial de iteraciones de las estimaciones de los parámetros y la log-verosimilitud, e imprime la última evaluación del vector de gradiente y la matriz hessiana. La tabla del historial de iteraciones muestra las estimaciones de los parámetros para cada *n* iteraciones a partir de la iteración 0 (las estimaciones iniciales), donde

*n* es el valor del intervalo de impresión. Si se solicita el historial de iteraciones, la última iteración siempre se muestra independientemente de *n*.

- **Matriz de correlaciones de trabajo.** Muestra los valores de la matriz que representan las dependencias intra-sujetos. Su estructura depende de las especificaciones de la pestaña Repetida.

## *Ecuaciones de estimación generalizadas: Medias marginales estimadas*

Figura 7-11
*Ecuaciones de estimación generalizadas: Pestaña Medias marginales estimadas*




Esta pestaña permite ver medias marginales estimadas para los niveles de factores y las interacciones de los factores. También se puede solicitar que se muestre la media estimada global. Las medias marginales estimadas no están disponibles para modelos multinomiales ordinales.

**Factores e interacciones.** Esta lista contiene los factores especificados en la pestaña Predictores y las interacciones de los factores especificadas en la pestaña Modelo. Las covariables se excluyen de esta lista. Los términos pueden seleccionar directamente en esta lista o combinarse en un término de interacción utilizando el botón Por *.

**Mostrar las medias para.** Se calculan las medias estimadas de los factores seleccionados y las interacciones de los factores. El contraste determina como se configuran los contrastes de hipótesis para comparar las medias estimadas. El contraste simple requiere una categoría de referencia o un nivel de factor con el que comparar los demás.

- **Por parejas.** Se calculan las comparaciones por parejas para todas las combinaciones de niveles de los factores especificados o implicados. Este contraste es el único disponible para las interacciones de los factores.
- **Simple.** Compara la media de cada nivel con la media de un nivel especificado. Este tipo de contraste resulta útil cuando existe un grupo de control.
- **Desviación.** Cada nivel del factor se compara con la media global. Los contrastes de desviación no son ortogonales.
- **Diferencia.** Compara la media de cada nivel (excepto el primero) con la media de los niveles anteriores. En ocasiones se les denomina contrastes de Helmert invertidos.
- **Helmert.** Compara la media de cada nivel del factor (excepto el último) con la media de los niveles siguientes.
- **Repetido.** Compara la media de cada nivel (excepto el último) con la media del nivel siguiente.
- **Polinómico.** Compara el efecto lineal, cuadrático, cúbico, etc. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, y así sucesivamente. Estos contrastes se utilizan a menudo para estimar las tendencias polinómicas.

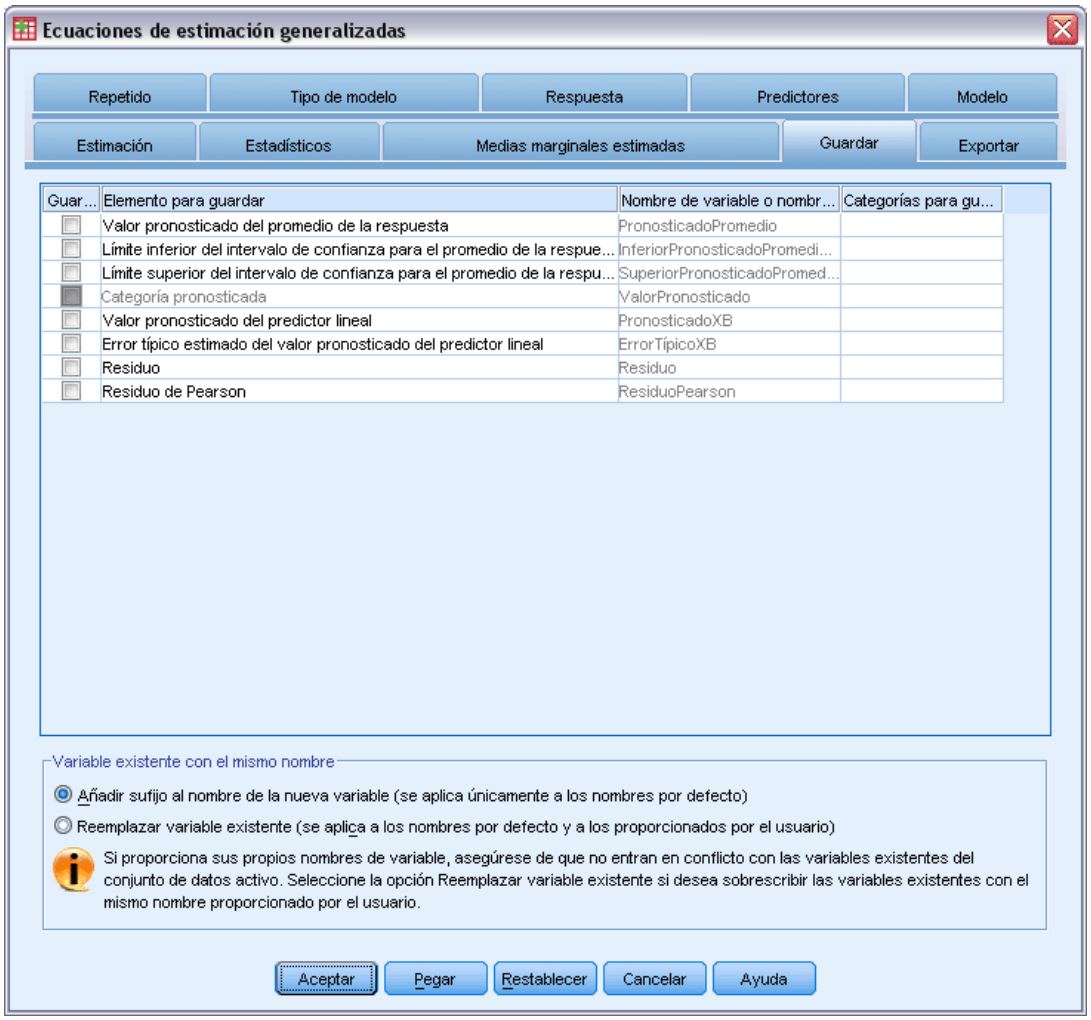
**Escala.** Se pueden calcular las medias marginales estimadas de la respuesta, basadas en la escala original de la variable dependiente o, para el predictor lineal, basadas en la variable dependiente tal como la transforma la función de enlace.

**Corrección para comparaciones múltiples.** Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Este grupo permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- **Bonferroni.** Este método corrige el nivel de significación observado por el hecho de que se están poniendo a prueba múltiples contrastes.
- **Bonferroni secuencial.** Este es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- **Sidak.** Este método ofrece límites más estrechos que los de la aproximación de Bonferroni.
- **Sidak secuencial.** Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.

# Ecuaciones de estimación generalizadas: Guardar

Figura 7-12
*Ecuaciones de estimación generalizadas: Pestaña Guardar*




Los elementos marcados se guardan con el nombre especificado. Puede elegir si desea sobrescribir las variables existentes con el mismo nombre que las nuevas variables o evitar conflictos de nombres adjuntando sufijos para asegurarse de que los nombres de las nuevas variables son únicos.

- **Valor pronosticado del promedio de la respuesta.** Guarda los valores pronosticados por el modelo para cada caso en la métrica de respuesta original. Cuando la distribución de la respuesta es binomial y la variable dependiente es binaria, el procedimiento guarda las probabilidades pronosticadas. Cuando la distribución de la respuesta es multinomial, la etiqueta del elemento se convierte en Probabilidad pronosticada acumulada y el procedimiento guarda la probabilidad pronosticada acumulada de cada categoría de la respuesta, salvo la última, hasta el número de categorías que se ha especificado que se guarden.

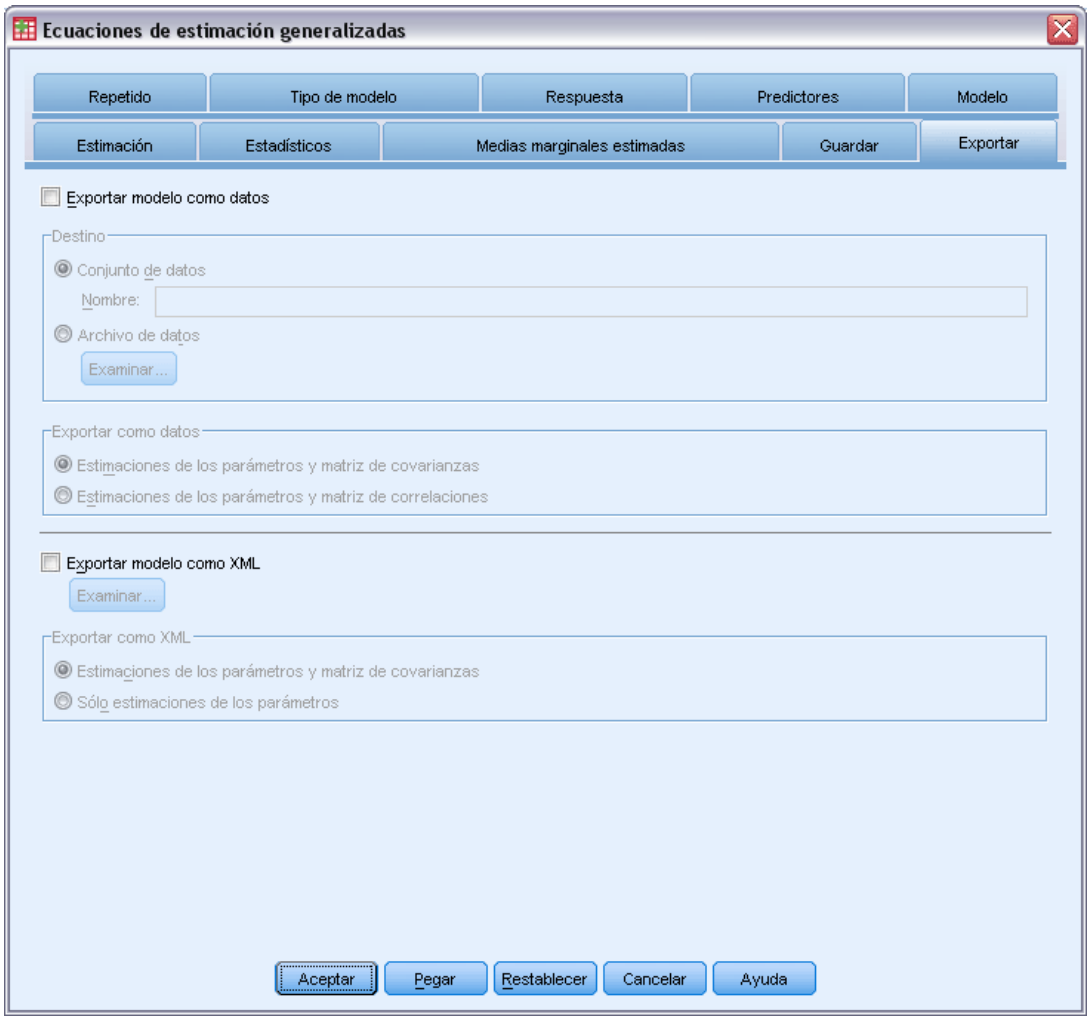
- **Límite inferior del intervalo de confianza para el promedio de la respuesta.** Guarda el límite inferior del intervalo de confianza de la media de la respuesta. Cuando la distribución de la respuesta es multinomial, la etiqueta del elemento se convierte en Límite inferior del intervalo de confianza para la probabilidad pronosticada acumulada y el procedimiento guarda el límite inferior de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.
- **Límite superior del intervalo de confianza para el promedio de la respuesta.** Guarda el límite superior del intervalo de confianza de la media de la respuesta. Cuando la distribución de la respuesta es multinomial, la etiqueta del elemento se convierte en Límite superior del intervalo de confianza para la probabilidad pronosticada acumulada y el procedimiento guarda el límite superior de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.
- **Categoría pronosticada.** Para los modelos con distribución binomial y variable dependiente binaria, o distribución multinomial, esta opción guarda la categoría de respuesta pronosticada para cada caso. Esta opción no esta disponible para otras distribuciones de la respuesta.
- **Valor pronosticado del predictor lineal.** Guarda los valores pronosticados por el modelo para cada caso en la métrica del predictor lineal (respuesta transformada mediante la función de enlace especificada). Cuando la distribución de respuesta es multinomial, el procedimiento guarda el valor pronosticado de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.
- **Error típico estimado del valor pronosticado del predictor lineal.** Cuando la distribución de respuesta es multinomial, el procedimiento guarda el error típico estimado de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.

Los siguientes elementos no están disponibles cuando la distribución de la respuesta es multinomial.

- **Residuo bruto.** Diferencia entre un valor observado y el valor pronosticado por el modelo.
- **Residuo de Pearson.** La raíz cuadrada de la contribución de un caso al estadístico chi-cuadrado de Pearson, con el signo del residuo bruto.

# Ecuaciones de estimación generalizadas: Exportar

Figura 7-13
*Ecuaciones de estimación generalizadas: pestaña Exportar*




**Exportar modelo como datos.** Escribe un conjunto de datos de formato IBM® SPSS® Statistics que contiene la matriz de covarianzas o correlaciones de los parámetros con las estimaciones de los parámetros, errores típicos, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **Variables de segmentación.** Si se han utilizado, todas las variables que definan segmentaciones.
- **RowType_.** Toma los valores (y las etiquetas de valor) *COV* (covarianzas), *CORR* (correlaciones), *EST* (estimaciones de los parámetros), *SE* (errores típicos), *SIG* (niveles de significación) y *DF* (grados de libertad del diseño muestral). Hay un caso diferente con el tipo de fila *COV* (o *CORR*) para cada parámetro del modelo, además de un caso diferente para cada uno de los otros tipos de filas.

- **VarName_.** Toma los valores *P1*, *P2*, ..., correspondientes a una lista ordenada de todos los parámetros estimados del modelo (salvo los parámetros binomiales negativos o de escala), para los tipos de fila *COV* o *CORR*, con las etiquetas de valor correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros. Las casillas están vacías para los demás tipos de filas.
- **P1, P2, ...** Estas variables corresponden a una lista ordenada de todos los parámetros del modelo (incluidos los parámetros binomiales negativos y de escala, según sea apropiado), con las etiquetas de variable correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros y toman valores según el tipo de fila.

  Para los parámetros redundantes, todas las covarianzas se establecen en cero, las correlaciones se establecen en el valor perdido del sistema; todas las estimaciones de los parámetros se establecen en cero; y todos los errores típicos, niveles de significación y los grados de libertad residuales se establecen en el valor perdido del sistema.

  Para el parámetro de escala, las covarianzas, correlaciones, nivel de significación y grados de libertad se establecen en el valor perdido del sistema. Si el parámetro de escala se estima mediante máxima verosimilitud, se indica el error típico; en otro caso se establece en el valor perdido del sistema.

  Para el parámetro binomial negativo, las covarianzas, correlaciones, nivel de significación y grados de libertad se establecen en el valor perdido del sistema. Si el parámetro binomial negativo se estima mediante máxima verosimilitud, se indica el error típico; en otro caso se establece en el valor perdido del sistema.

  Si hay segmentaciones, se debe acumular la lista de parámetros a través de todas las segmentaciones. En una determinada segmentación, es posible que algunos parámetros sean irrelevantes; pero no es lo mismo que sean redundantes. Para los parámetros irrelevantes, todas las covarianzas y correlaciones, estimaciones de los parámetros, errores típicos, niveles de significación y grados de libertad se establecen en el valor perdido del sistema.

Puede utilizar este archivo matricial como valores iniciales para una estimación posterior del modelo; tenga en cuenta que este archivo no se puede utilizar directamente para realizar otros análisis en otros procedimientos que lean un archivo matricial a menos que dichos procedimientos acepten todos los tipos de filas que aquí se exportan. Incluso en esos casos, deberá asegurarse de que todos los parámetros del archivo matricial tienen el mismo significado para el procedimiento que lee el archivo.

**Exportar modelo como XML.** Guarda las estimaciones de los parámetros y la matriz de covarianzas de los parámetros (si se selecciona) 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 GENLIN

La sintaxis de comandos también le permite:

- Especificar valores iniciales para las estimaciones de los parámetros como una lista de números (utilizando el subcomando `CRITERIA`).
- Especificar una matriz de correlaciones de trabajo fija (utilizando el subcomando `REPEATED`).

- Fijar covariables en valores distintos los de sus medias al calcular las medias marginales estimadas (utilizando el subcomando `EMMEANS`).
- Especificar contrastes polinómicos personalizados para las medias marginales estimadas (utilizando el subcomando `EMMEANS`).
- Especificar un subconjunto de los factores para los que se muestran las medias marginales estimadas para compararlos utilizando el tipo de contraste especificado (utilizando las palabras clave `TABLES` y `COMPARE` del subcomando `EMMEANS`).

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

Capítulo
8

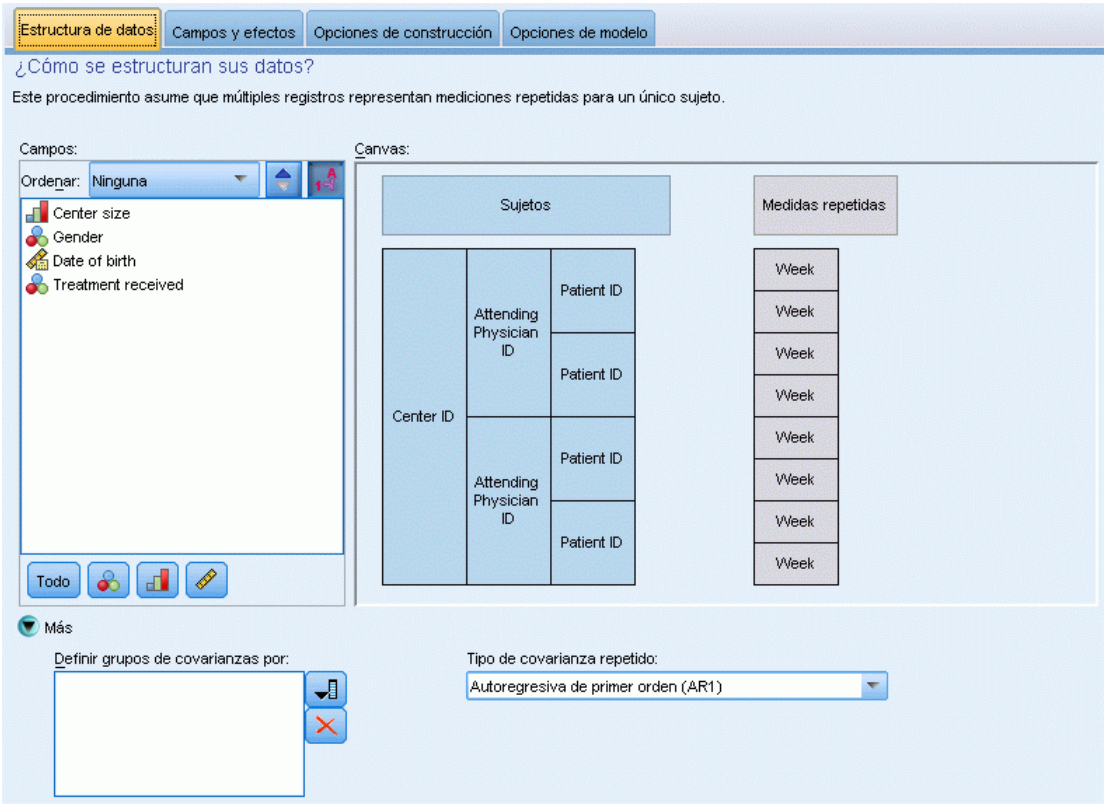
# Modelos mixtos lineales generalizados

Los modelos mixtos lineales generalizados amplían el modelo lineal de forma que:

- El destino tenga una relación lineal con los factores y covariables mediante una función de enlace especificada.
- El destino pueda tener una distribución no normal.
- Las observaciones pueden estar correlacionadas.

Los modelos mixtos lineales generalizados cubren una amplia variedad de modelos, desde modelos mixtos lineales generalizados a modelos multinivel complejos de datos longitudinales no normales.

Figura 8-1
*Pestaña Estructura de datos*



La pestaña Estructura de datos le permite especificar las relaciones estructurales entre los registros de su conjunto de datos cuando se correlacionan las observaciones. Si los registros del conjunto de datos representan observaciones independientes, no deberá especificar nada en esta pestaña.

**Sujetos.** La combinación de valores de los campos categóricos especificados debe definir de manera única los sujetos del conjunto de datos. Por ejemplo, un campo único *ID de paciente* debería ser suficiente para definir los sujetos de un único hospital, pero puede que sea necesario combinar *ID de hospital* e *ID de paciente* si los números de identificación de paciente no son únicos entre varios hospitales. En una configuración de medidas repetidas, se registran varias observaciones para cada sujeto, de manera que cada sujeto puede ocupar varios registros del conjunto de datos.

Un **sujeto** es una unidad de observación, la cual se puede considerar independiente de otros sujetos. Por ejemplo, en un estudio médico, las lecturas de la presión sanguínea de un paciente se pueden considerar independientes de las lecturas de otros pacientes. La definición de los sujetos es particularmente importante cuando se dan medidas repetidas para cada sujeto y desea modelar la correlación entre estas observaciones. Por ejemplo, cabe esperar que estén correlacionadas las lecturas de la presión sanguínea de un único paciente en una serie de visitas consecutivas al médico.

Todos los campos especificados como Sujetos en la pestaña Estructura de datos se utilizan para definir sujetos para la estructura de la covarianza residual y obtener la lista de posibles campos para definir sujetos para estructuras de covarianza de los efectos aleatorios en el Bloque de efectos aleatorios.

**Medidas repetidas.** Los campos especificados aquí se usan para identificar las observaciones repetidas. Por ejemplo, una única variable *Semana* puede identificar las 10 semanas de observaciones de un estudio médico o se pueden usar *Mes* y *Día* para identificar las observaciones diarias realizadas a lo largo de un año.

**Definir grupos de covarianza por.** Los campos que se especifiquen aquí definen conjuntos independientes de parámetros de covarianza de efectos repetidos, uno por cada categoría definida por la clasificación cruzada de los campos de agrupación. Todos los sujetos tienen el mismo tipo de covarianza; los sujetos en el mismo grupo de covarianza tendrán los mismos valores de los parámetros.

**Tipo de covarianza para Repetidas.** Especifica la estructura de la covarianza para los residuos. Las estructuras disponibles son:

- Autorregresiva de primer orden (AR1)
- Media móvil autorregresiva (1,1) (ARMA11)
- Simetría compuesta
- Diagonal
- Identidad escalada
- Toeplitz
- Sin estructura
- Componentes de la varianza

Si desea obtener más información, consulte el tema Estructuras de covarianza en el apéndice B el p. 166.

# Obtención de un modelo mixto lineal generalizado

Esta función requiere la opción Estadísticas avanzadas.

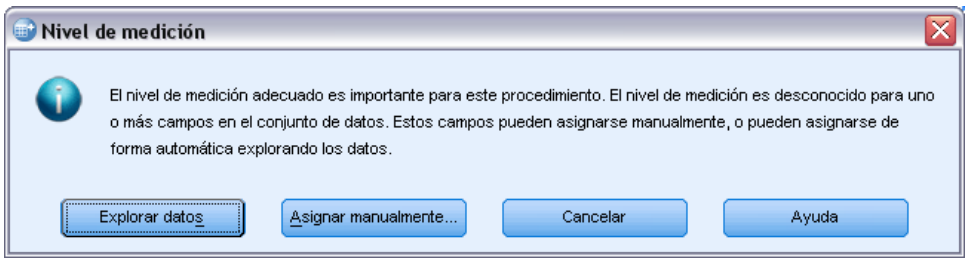
Seleccione en los menús:
Analizar > Modelos mixtos > Lineal generalizado...

- Defina la estructura del sujeto de su conjunto de datos en la pestaña Estructura de datos.
- En la pestaña Campos y efectos debe haber un único destino que puede tener cualquier nivel de medición o una especificación de eventos/ensayos, en cuyo caso las especificaciones de eventos y ensayos deben ser continuas. También puede especificar su distribución y función de enlace, los efectos fijos y cualquier bloque de efectos aleatorios, desplazamiento o ponderaciones de análisis.
- Pulse en Opciones de generación para especificar cualquier configuración de generación.
- Pulse en Opciones de modelo para guardar puntuaciones en el conjunto de datos activo y exportar el modelo en un archivo externo.
- Pulse en Ejecutar para ejecutar el procedimiento y crear los objetos Modelo.

### *Campos con un nivel de medición desconocido*

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 8-2
*Alerta de nivel de medición*

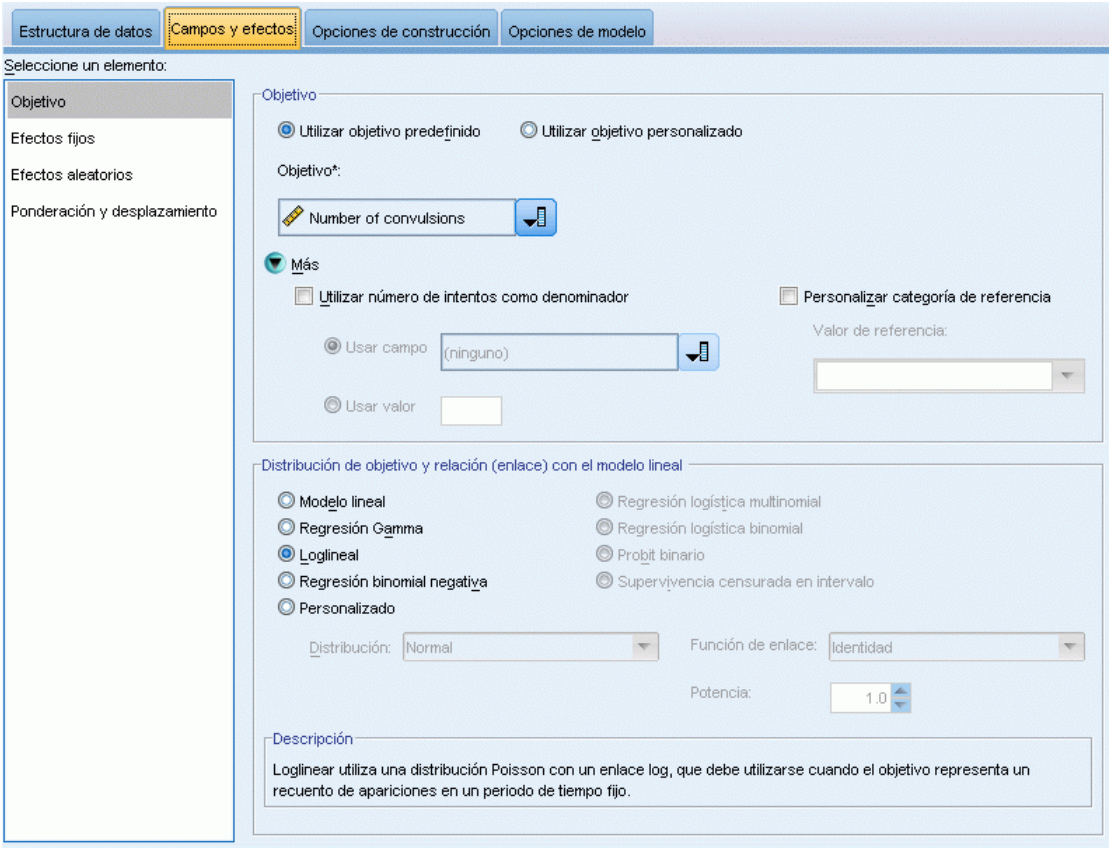



- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

# *Objetivo*

Figura 8-3
*Configuración de destino*




Esta configuración define el destino, su distribución y su relación con los predictores mediante la función de enlace.

**Objetivo.** El objetivo es obligatorio. Puede tener cualquier nivel de medición y el nivel de medición del destino restringe las distribuciones y funciones de enlace que son adecuadas.

- **Use el número de ensayos como denominador.** Cuando la respuesta de destino es un número de eventos que ocurren en un conjunto de ensayos, el campo de destino contiene el número de eventos y puede seleccionar una variable adicional que contenga el número de ensayos. Por ejemplo, al probar un nuevo pesticida podría exponer muestras de hormigas a diferentes concentraciones del pesticida y después registrar el número de hormigas muertas y el número de hormigas de cada muestra. En este caso, el campo que registra el número de hormigas muertas debe especificarse como el campo de destino (eventos), y el que registra el número de hormigas de cada muestra debe especificarse como el campo de ensayos. Si el número

de hormigas es el mismo para cada muestra, el número de ensayos podrá especificarse usando un valor fijo.

El número de ensayos debe ser mayor o igual que el número de eventos para cada registro. Los eventos deben ser enteros no negativos y los ensayos deben ser enteros positivos.

- **Personalizar la categoría de referencia.** Para un destino categórico, puede seleccionar la categoría de referencia. Esto puede afectar a ciertos resultados, como las estimaciones de los parámetros, pero no debería cambiar el ajuste del modelo. Por ejemplo, si su destino toma los valores 0, 1 y 2, por defecto el procedimiento realiza la última categoría (la de mayor valor) o 2, la categoría de referencia. En esta situación, las estimaciones de parámetros deben interpretarse como relacionadas con la probabilidad de la categoría 0 ó 1 *relativa* a la probabilidad de que haya una categoría 2. Si especifica una categoría personalizada y su destino tiene etiquetas definidas, puede definir la categoría de referencia seleccionando un valor de la lista. Puede resultar cómodo si, a mitad del proceso de especificar un modelo, no recuerda exactamente cómo se ha codificado un campo concreto.

**Distribución de destino y relación (enlace) con el modelo lineal.** Teniendo en cuenta los valores de los predictores, el modelo espera que la distribución de los valores del destino siga la forma especificada y que los valores de destino tengan una relación lineal con los predictores mediante la función de enlace especificada. Se proporcionan los accesos directos de varios modelos comunes o seleccione un ajuste Personalizado si hay una combinación específica de distribución y función de enlace que desee ajustar y que no esté en la lista corta.

- **Modelo lineal.** Especifica una distribución normal con un enlace de identidad, que es útil si el destino se puede pronosticar mediante una regresión lineal o un modelo ANOVA.
- **Regresión gamma.** Especifica una distribución gamma con un enlace de logaritmo, que se debe utilizar si el destino contiene todos los valores positivos y es asimétrico a valores mayores.
- **Loglinear.** Especifica una distribución de Poisson con un enlace de logaritmo, que se debe utilizar si el destino representa un recuento de apariciones en un periodo de tiempo fijo.
- **Regresión binomial negativa.** Especifica una distribución binomial negativa con un enlace de logaritmo, que se debe utilizar si el destino y el denominador representan el número de ensayos necesarios para observar *k* éxitos.
- **Regresión logística multinomial.** Especifica una distribución multinomial con un enlace Logit generalizado, que se debe utilizar si el destino es una respuesta de categoría múltiple.
- **Regresión logística binaria.** Especifica una distribución binomial con un enlace Logit, que se debe utilizar si el destino es una respuesta binaria pronosticada por un modelo de regresión logística.
- **Probit binario.** Especifica una distribución binomial con un enlace probit, que se debe utilizar si el destino es una respuesta binaria con una distribución normal subyacente.
- **Supervivencia censurada en intervalo.** Especifica una distribución binomial con un enlace log-log complementario, que es útil en un análisis de supervivencia si algunas observaciones no tienen eventos de finalización.

### *Distribución*

Esta selección especifica la distribución del destino. La posibilidad de especificar una distribución que no sea la normal y una función de enlace que no sea la identidad es la principal mejora que aporta el modelo mixto lineal generalizado respecto al modelo lineal general. Hay muchas combinaciones posibles de distribución y función de enlace, varias de las cuales pueden ser adecuadas para un determinado conjunto de datos, por lo que su elección puede estar guiada por consideraciones teóricas a priori y por las combinaciones que parezcan funcionar mejor.

- **Binomial.** Esta distribución es adecuada únicamente para un destino que represente una respuesta binaria o un número de eventos.
- **Gamma.** Esta distribución es adecuada para un destino con valores de escala positivos que se desvían hacia valores positivos más grandes. Si un valor de datos es menor o igual que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **De Gauss inversa.** Esta distribución es adecuada para un destino con valores de escala positivos que se desvían hacia valores positivos más grandes. Si un valor de datos es menor o igual que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.
- **Multinomial.** Esta distribución es adecuada para un destino que represente una respuesta de categoría múltiple.
- **Binomial negativa.** La regresión binomial negativa utiliza una distribución binomial negativa con un enlace de logaritmo, que se debe utilizar si el destino representa un recuento de apariciones con una alta varianza.
- **Normal.** Es adecuada para un destino continuo cuyos valores adoptan una distribución simétrica con forma de campana en torno a un valor central (la media).
- **Poisson.** Esta distribución considera el número de ocurrencias de un evento de interés en un período fijo de tiempo y es apropiada para variables que tengan valores enteros que no sean negativos. Si un valor de datos no es entero, es menor que 0 o es un valor perdido, el correspondiente caso no se utilizará en el análisis.

### *Funciones de enlace*

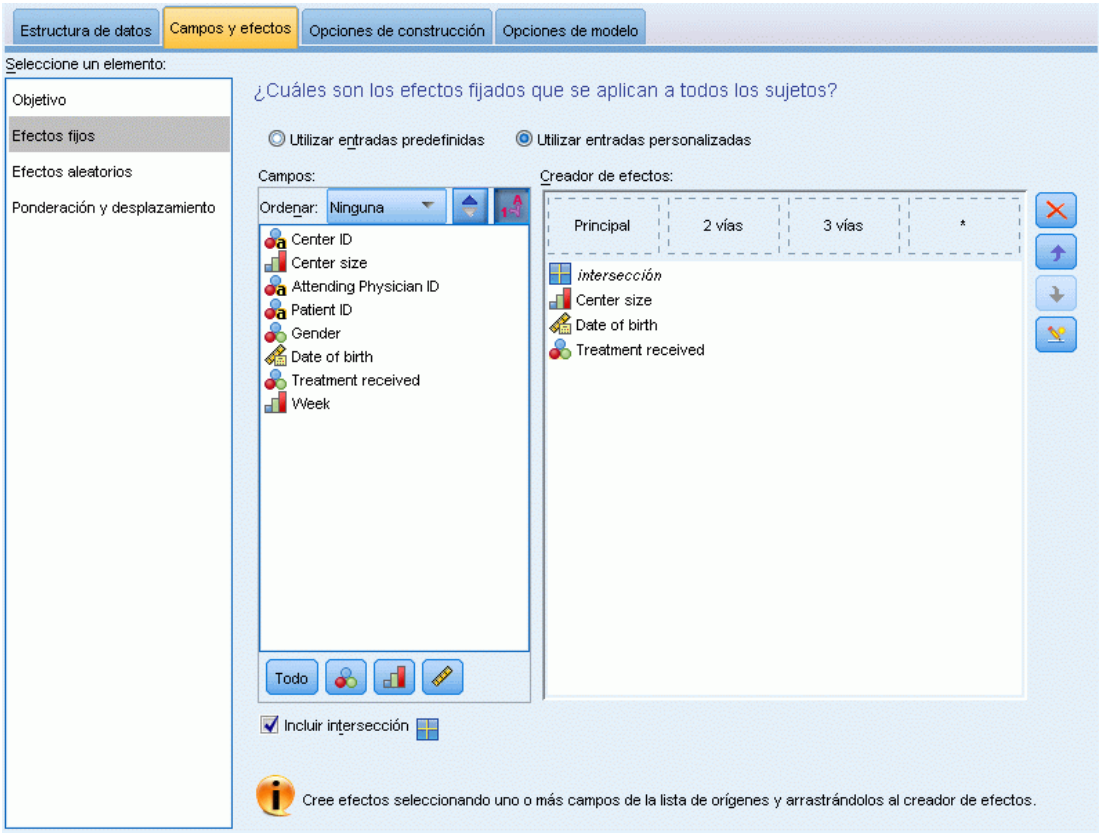
La función de enlace es una transformación del destino que permite la estimación del modelo. Se encuentran disponibles las siguientes funciones:

- **Identidad.** $f(x)=x$. El destino no se transforma. Este vínculo se puede utilizar con cualquier distribución, excepto la multinomial.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Es apropiada únicamente para la distribución binomial.
- **Log.** $f(x)=\log(x)$. Este vínculo se puede utilizar con cualquier distribución, excepto la multinomial.
- **Complemento log.** $f(x)=\log(1-x)$. Es apropiada únicamente para la distribución binomial.
- **Logit.** $f(x)=\log(x / (1-x))$. Esto sólo es apropiado con las distribuciones binomiales o multinomiales, ya que es la única función de enlace válida para el multinomial.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Es apropiada únicamente para la distribución binomial.

- **Probit.** $f(x)=\Phi^{-1}(x)$, donde $\Phi^{-1}$ es la función de distribución acumulada normal típica inversa. Es apropiada únicamente para la distribución binomial.
- **Potencia.** $f(x)=x^{\alpha}$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. $\alpha$ es la especificación de número necesaria y debe ser un número real. Este vínculo se puede utilizar con cualquier distribución, excepto la multinomial.

# Efectos fijos

Figura 8-4
*Configuración de efectos fijos*




Los factores de efectos fijos se suelen considerar campos cuyos valores de interés se representan en el conjunto de datos y se pueden utilizar para la puntuación. Por defecto, los campos con el papel de entrada predefinido que no se especifican en ninguna otra parte del cuadro de diálogo se introducen en la sección de efectos fijos del modelo. Los campos categóricos (nominal y ordinal) se utilizan como factores en el modelo y los campos continuos se utilizan como covariables.

Introduzca los efectos en el modelo seleccionando uno o más campos en la lista de origen arrastrando la lista de efectos. El tipo de efecto que se crea depende de la zona activa en la que suelte la selección.

- **Principal.** Los campos aparecen como efectos principales diferentes en la parte inferior de la lista de efectos.

- **2 vías.** Todos los pares posibles de los campos aparecerán como interacciones de 2 vías en la parte inferior de la lista de efectos.
- **3 vías.** Todos los triples posibles de los campos aparecerán como interacciones de 3 vías en la parte inferior de la lista de efectos.
- ***.** La combinación de todos los campos aparecerá como una interacción única en la parte inferior de la lista de efectos.

Los botones a la derecha del creador de efectos le permiten:

Eliminar términos del modelo de efectos fijos seleccionando los que desea eliminar y pulsando en el botón eliminar.

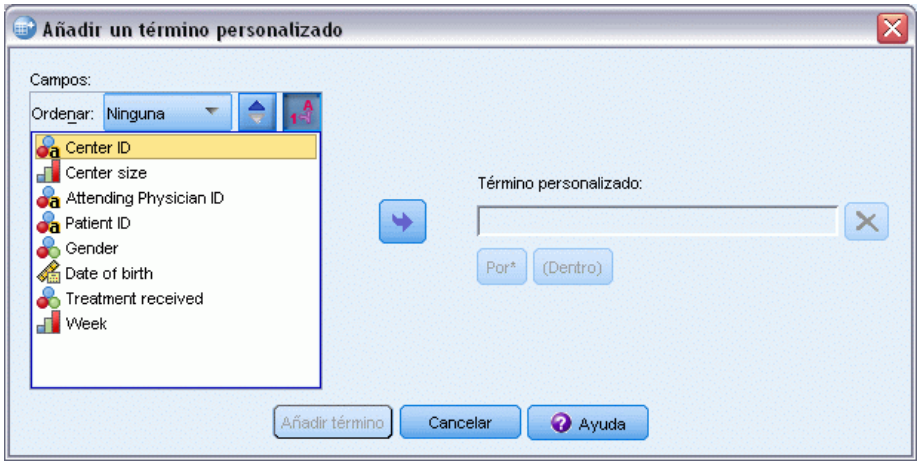
Volver a ordenar los términos del modelo de efectos fijos seleccionando los que desea volver a ordenar y pulsando en la flecha arriba o abajo.

Añadir términos anidados al modelo usando el cuadro de diálogo Añadir un término personalizado , pulsando en el botón Añadir un término personalizado.

**Incluir intersección.** La intersección se incluye normalmente en el modelo. Si asume que los datos pasan por el origen, puede excluir la intersección.

## *Añadir un término personalizado*

Figura 8-5
*Cuadro de diálogo Añadir un término personalizado*




En este procedimiento, puede construir términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción, como términos polinómicos que implican a la misma covariable, o añadir varios niveles de anidación al término anidado.

**Limitaciones.** Existen las siguientes restricciones para los términos anidados:

- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A*A*.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si *A* es un factor, no es válido especificar *A(A)*.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si *A* es un factor y *X* es una covariable, no es válido especificar *A(X)*.

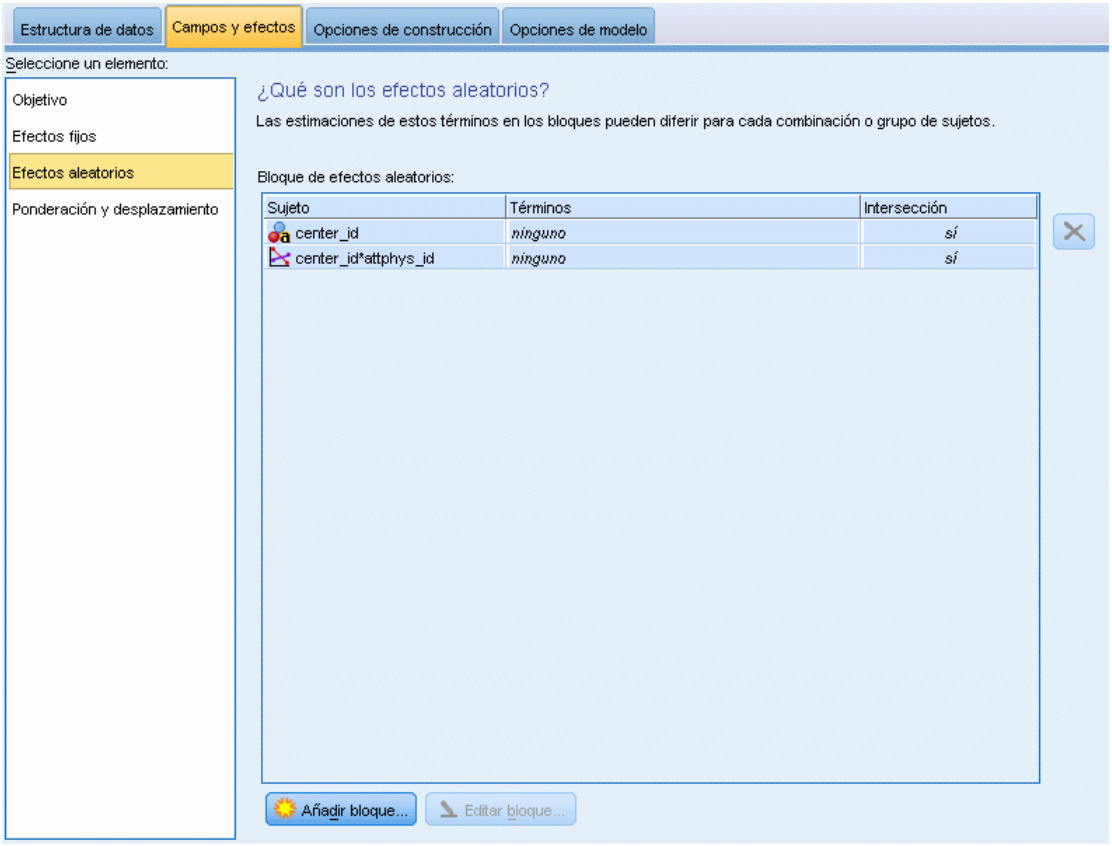
***Construcción de un término anidado***

- Seleccione un factor o covariable que esté anidado en otro factor y, a continuación, pulse en el botón de flecha.
- Pulse en (Dentro).
- Seleccione el factor dentro del cual el factor o covariable anterior se anida y pulse en el botón de flecha.
- Pulse en Añadir término.

Si lo desea, puede incluir efectos de interacción o añadir varios niveles de anidación al término anidado.

# Efectos aleatorios

Figura 8-6
*Configuración de efectos aleatorios*




Los factores de efectos aleatorios son campos cuyos valores en el archivo de datos se pueden considerar una muestra aleatoria de una población mayor de valores. Son útiles para explicar el exceso de variabilidad en el destino. Por defecto, si ha seleccionado más de un sujeto en la pestaña Estructura de datos, se creará un bloque de efectos aleatorios para cada sujeto más allá de su sujeto más interior. Por ejemplo, si ha seleccionado Colegio, Clase y Alumno como sujetos en la pestaña Estructura de datos, se crearán automáticamente los siguientes bloques de efectos aleatorios:

- Efecto aleatorio 1: el sujeto es colegio (sin efectos, intersección solamente)
- Efecto aleatorio 2: el sujeto es colegio * clase (sin efectos, intersección solamente)

Puede trabajar con bloques de efectos aleatorios de las siguientes formas:

- Para añadir un bloque nuevo, pulse en Añadir bloque...Se abrirá el cuadro de diálogo Bloque de efectos aleatorios .
- Para modificar un bloque existente, seleccione el bloque que desea modificar y pulse en Editar bloque... Se abrirá el cuadro de diálogo Bloque de efectos aleatorios .
- Para eliminar uno o más bloques, seleccione los bloques que desee eliminar y pulse en botón Eliminar.

## *Bloque de efectos aleatorios*

Figura 8-7
*Cuadro de diálogo Bloque de efectos aleatorios*


Introduzca los efectos en el modelo seleccionando uno o más campos en la lista de origen arrastrando la lista de efectos. El tipo de efecto que se crea depende de la zona activa en la que suelte la selección. Los campos categóricos (nominal y ordinal) se utilizan como factores en el modelo y los campos continuos se utilizan como covariables.

- **Principal.** Los campos aparecen como efectos principales diferentes en la parte inferior de la lista de efectos.
- **2 vías.** Todos los pares posibles de los campos aparecerán como interacciones de 2 vías en la parte inferior de la lista de efectos.
- **3 vías.** Todos los triples posibles de los campos aparecerán como interacciones de 3 vías en la parte inferior de la lista de efectos.
- ***.** La combinación de todos los campos aparecerá como una interacción única en la parte inferior de la lista de efectos.

Los botones a la derecha del creador de efectos le permiten:

Eliminar términos del modelo de efectos fijos seleccionando los que desea eliminar y pulsando en el botón eliminar.

Volver a ordenar los términos del modelo de efectos fijos seleccionando los que desea volver a ordenar y pulsando en la flecha arriba o abajo.

Añadir términos anidados al modelo usando el cuadro de diálogo Añadir un término personalizado , pulsando en el botón Añadir un término personalizado.

**Incluir intersección.** La intersección se incluye normalmente en el modelo de efectos aleatorios de forma predeterminada. Si asume que los datos pasan por el origen, puede excluir la intersección.

**Definir grupos de covarianza por.** Los campos que se especifiquen aquí definen conjuntos independientes de parámetros de covarianza de efectos aleatorios, uno por cada categoría definida por la clasificación cruzada de los campos de agrupación. Es posible especificar un conjunto distinto de campos de agrupación para cada bloque de efectos aleatorios. Todos los sujetos tienen el mismo tipo de covarianza; los sujetos en el mismo grupo de covarianza tendrán los mismos valores de los parámetros.

**Combinación de sujetos.** Le permite especificar sujetos de efectos aleatorios desde combinaciones predefinidas de sujetos de la pestaña Estructura de datos. Por ejemplo, si *Colegio*, *Clase* y *Alumno* se definen como sujetos en la pestaña Estructura de datos y en ese orden, la lista desplegable Combinación de sujetos tendrá las opciones Ninguno, Colegio, Colegio * Clase y Colegio * Clase * Alumno.

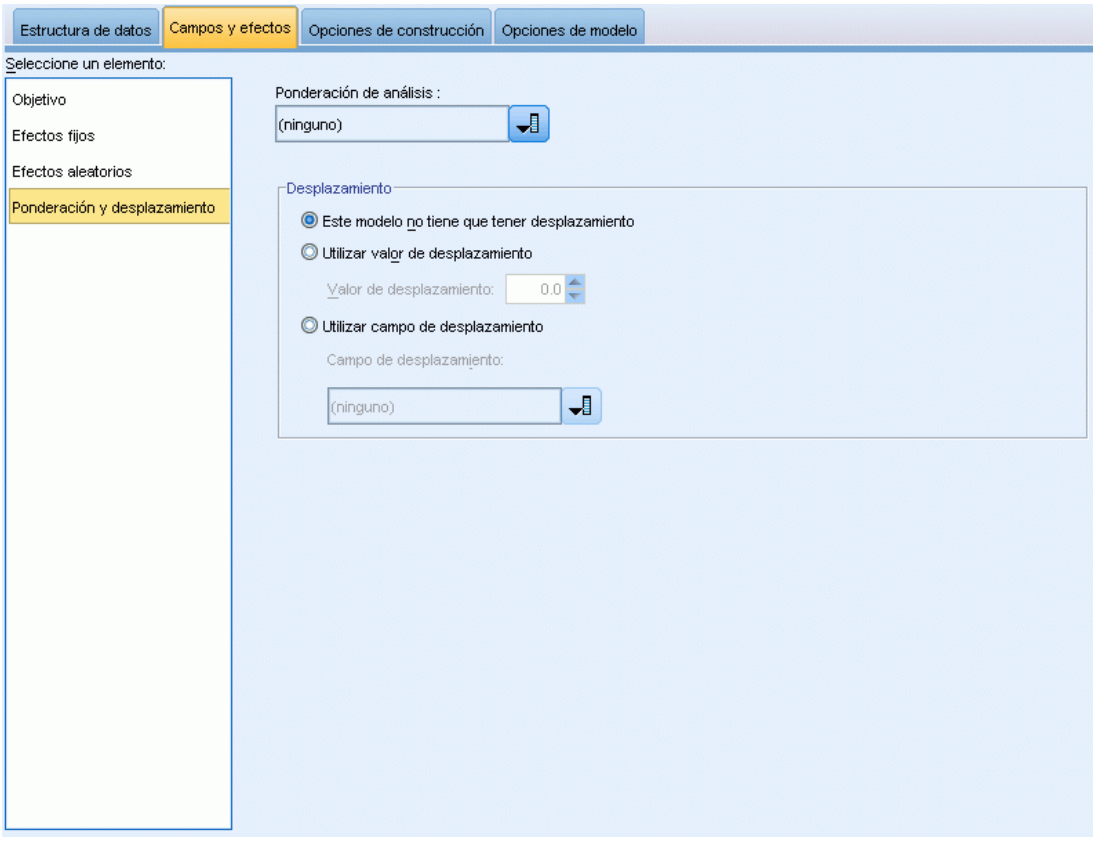
**Tipo de covarianza de efecto aleatorio.** Especifica la estructura de la covarianza para los residuos. Las estructuras disponibles son:

- Autorregresiva de primer orden (AR1)
- Media móvil autorregresiva (1,1) (ARMA11)
- Simetría compuesta
- Diagonal
- Identidad escalada
- Toeplitz
- Sin estructura
- Componentes de la varianza

Si desea obtener más información, consulte el tema Estructuras de covarianza en el apéndice B el p. 166.

# Ponderación y desplazamiento

Figura 8-8
*Configuración de Ponderación y desplazamiento*

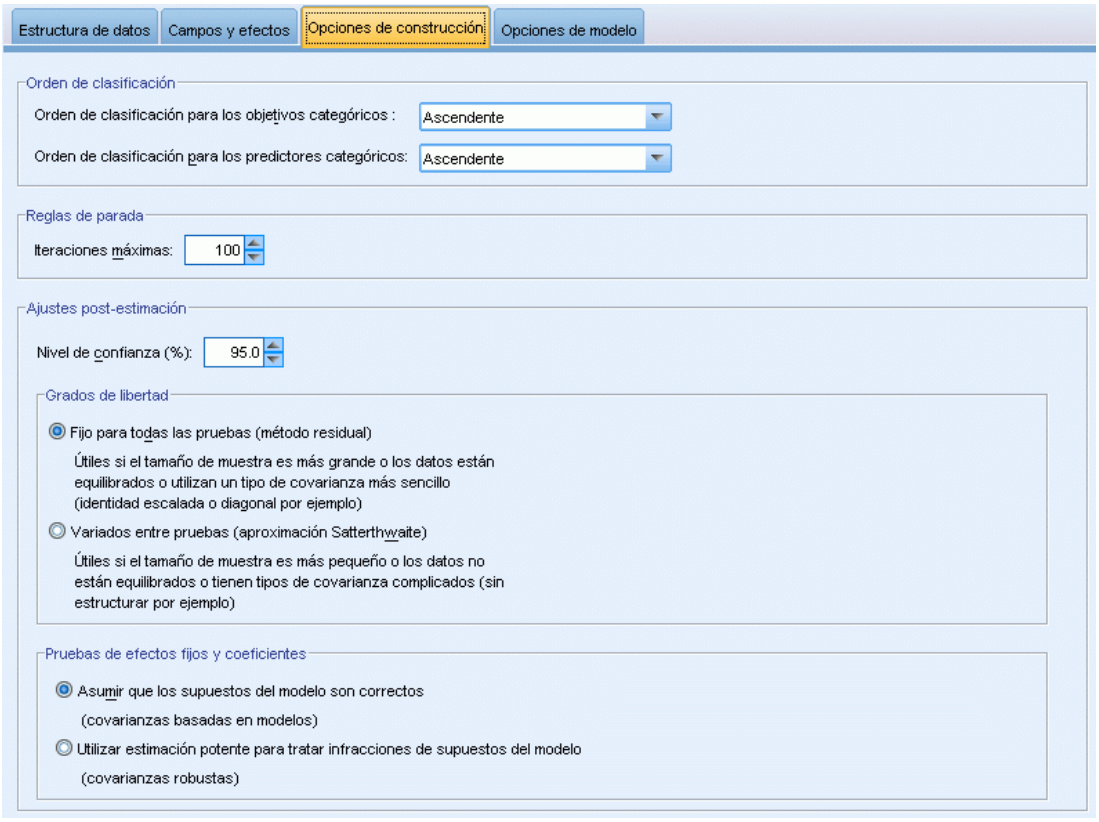



**Ponderación de análisis.** El parámetro de escala es un parámetro del modelo estimado relacionado con la varianza de la respuesta. Los pesos de análisis son valores "conocidos" que pueden variar de una observación a otra. Si se especifica el campo de ponderación de análisis, el parámetro de escala, que está relacionado con la varianza de la respuesta, se divide por los valores de ponderación de análisis para cada observación. Los registros cuyos valores de ponderación de análisis es menor o igual que 0 o que son perdidos no se utilizan en el análisis.

**Desplazamiento.** El término desplazamiento es un predictor "estructural". El modelo no estima su coeficiente, pero se supone que tiene el valor 1. Por tanto, los valores del desplazamiento se suman sencillamente al predictor lineal del destino. Esto resulta especialmente útil en los modelos de regresión de Poisson, en los que cada caso puede tener diferentes niveles de exposición al evento de interés. Por ejemplo, al modelar las tasas de accidente de diferentes conductores, hay una importante diferencia entre un conductor que ha sido el culpable de un accidente en tres años y un conductor que ha sido el culpable de un accidente en 25 años. El número de accidentes se puede modelar como una respuesta de Poisson si la experiencia del conductor se incluye como un término de desplazamiento.

# Opciones de construcción

Figura 8-9
*Configuración de Opciones de construcción*




Estas selecciones especifican algunos de los criterios más avanzados utilizados para crear el modelo.

**Orden de clasificación.** Estos controles determinan el orden de las categorías del destino y los factores (entradas categóricas) para determinar la "última" categoría. La configuración del orden de clasificación se ignora si el destino no es categórico o si se especifica una categoría de referencia personalizada en la configuración de Objetivo .

**Reglas de parada.** Puede especificar el número máximo de iteraciones que ejecutará el algoritmo. Especifique un número entero no negativo. El valor por defecto es 100.

**Ajustes post-estimación.** Esta configuración determina cómo se calculan algunos resultados del modelo para su visualización.

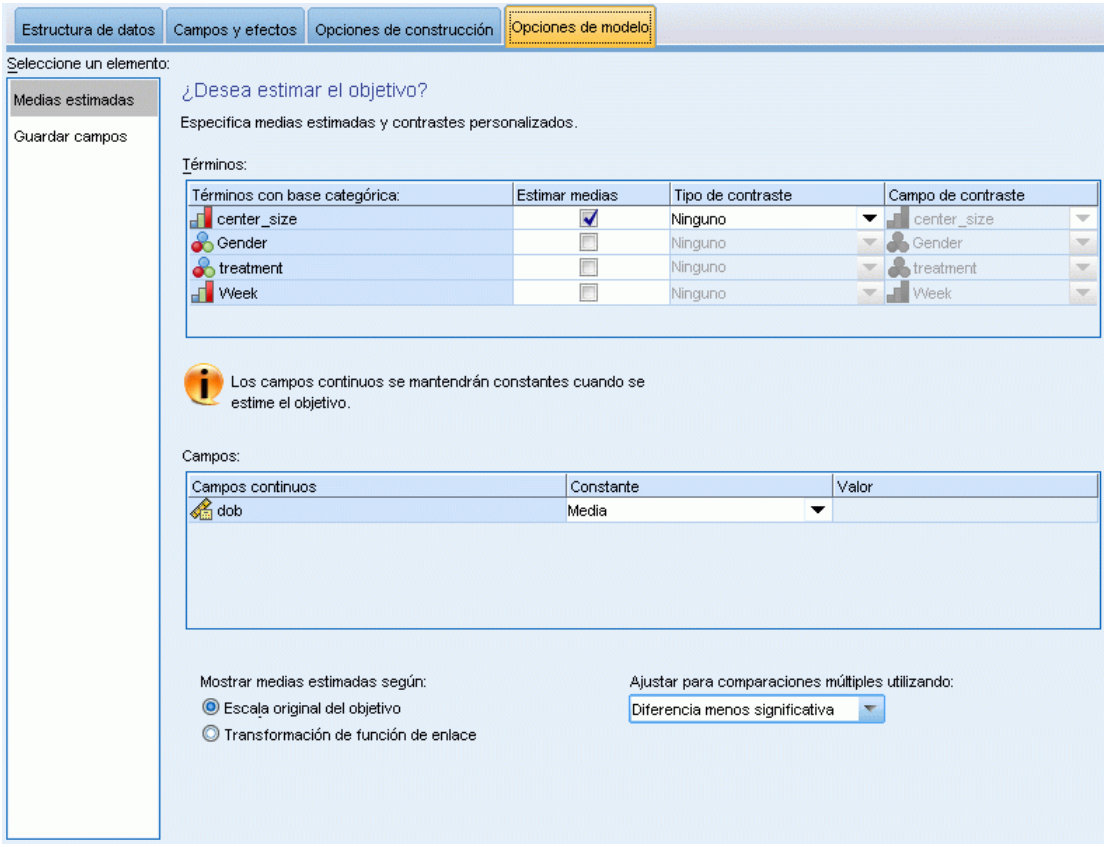
- **Nivel de confianza.** Éste es el nivel de confianza que se utiliza para calcular las estimaciones de intervalos de los coeficientes de modelos. Especifique un valor mayor que 0 y menor que 100. El valor por defecto es 95.
- **Grados de libertad.** Especifica cómo se calculan los grados de libertad para las comprobaciones de significación. Seleccione Fijo para todas las pruebas (método residual) si el tamaño de muestra es suficientemente grande, si datos están equilibrados o si el modelo utiliza un tipo

de covarianza más sencillo; por ejemplo, identidad escalada o diagonal. Ésta es la opción por defecto. Seleccione Variados entre pruebas (aproximación Satterthwaite) si el tamaño de muestra es pequeño, los datos no están equilibrados o el modelo utiliza un tipo de covarianza complicado; por ejemplo, sin estructurar.

- **Pruebas de efectos fijos y coeficientes.** Es el método para calcular la matriz de covarianzas de las estimaciones de los parámetros. Seleccione la estimación robusta si está preocupado porque se incumplan los supuestos del modelo.

# Medias estimadas

Figura 8-10
*Configuración de Medias estimadas*




Esta pestaña permite ver medias marginales estimadas para los niveles de factores y las interacciones de los factores. Las medias marginales estimadas no están disponibles para modelos multinomiales.

**Términos.** Los términos de modelo de los Efectos fijos que se componen enteramente de campos categóricos se enumeran aquí. Seleccione cada término para el que desea que el modelo produzca las medias marginales.

- **Tipo de contraste.** Especifica el tipo de contraste que se utilizará para los niveles del campo de contraste. Si selecciona Ninguno, no se producen contrastes. Por parejas produce comparaciones por parejas de todas las combinaciones de niveles de los factores especificados. Este contraste es el único disponible para las interacciones de los factores. Desviación compara cada nivel del factor con la media global. Los contrastes simples comparan cada nivel del factor, excepto el último, con el último nivel. El "último" nivel está determinado por el orden de clasificación de los factores especificados en Opciones de construcción. Tenga en cuenta que todos estos tipos de contrastes no son ortogonales.
- **Campo de contraste.** Especifica un factor, cuyos niveles se comparan utilizando el tipo de contraste seleccionado. Si se selecciona Ninguno como tipo de contraste, no podrá (ni será necesario) seleccionar ningún campo de contraste.

**Campos continuos.** Los campos continuos enumerados se extraen de los términos de los efectos fijos que usan campos continuos. Al calcular las medias marginales, las covariables se fijan en los valores especificados. Seleccione la media o especifique un valor personalizado.

**Mostrar medias estimadas según.** Especifica si se calcularán las medias marginales en función de la escala original del destino o en función de la transformación de la función de enlace. Escala original del objetivo calcula las medias marginales del destino. Tenga en cuenta que si se especifica el destino utilizando la opción eventos/ensayos, proporciona la media marginal para la proporción eventos/ensayos en lugar del número de eventos. Transformación de función de enlace calcula la media marginal del predictor lineal.

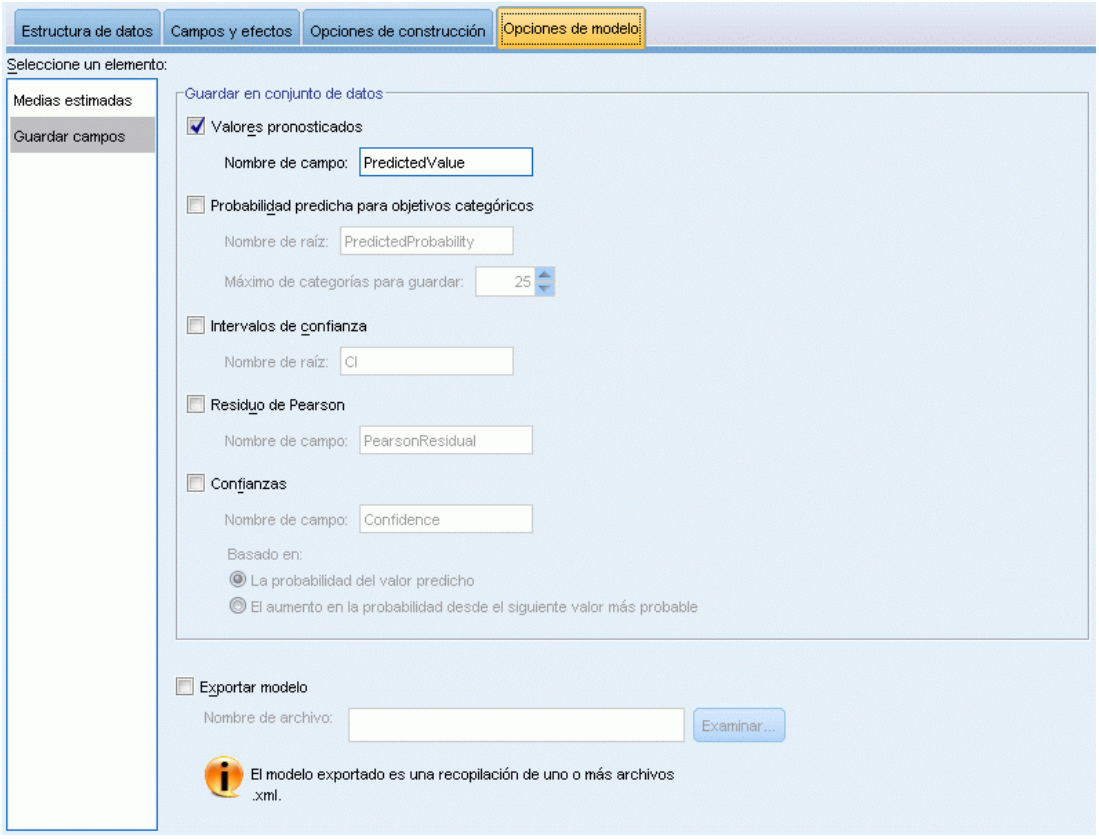
**Ajustar para comparaciones múltiples utilizando.** Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- **Bonferroni secuencial.** Este es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- **Sidak secuencial.** Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.

El método de diferencia menos significativa es menos conservador que el método de Sidak secuencial, que a su vez es menos conservador que el de Bonferroni secuencial; en otras palabras, el método de diferencia menos significativa rechazará como mínimo tantas hipótesis como el método de Sidak secuencial, que a su vez rechazará como mínimo tantas hipótesis como el método de Bonferroni secuencial.

# Guardar

Figura 8-11
*Configuración de guardar*




Los elementos seleccionados se guardan con el nombre especificado; no se permiten conflictos con los nombres del campo existente.

**Valores pronosticados.** Guarda el valor pronosticado del destino. El nombre del campo por defecto es *PredictedValue*.

**Probabilidad predicha de destinos categóricos.** Si el destino es categórico, esta palabra clave guarda las probabilidades pronosticadas de las primeras *n* categorías, hasta el valor especificado como Máximo de categorías para guardar. El nombre de raíz por defecto es *PredictedProbability*. Para guardar la probabilidad pronosticada de la categoría pronosticada, guarde la confianza (consulte a continuación).

**Intervalos de confianza.** Guarda el límite inferior y superior del intervalo de confianza del valor pronosticado o la probabilidad pronosticada. Para todas las distribuciones excepto la multinomial, crea dos variables y el nombre de raíz por defecto es *CI*, con *_Lower* y *_Upper* como sufijos.

Para la distribución multinomial, se crea un campo para cada categoría de variable dependiente, excepto la última (Si desea obtener más información, consulte el tema Opciones de construcción el p. 113.). Guarda los límites inferior y superior de la probabilidad acumulada pronosticada para las

*n* primeras categorías, hasta la última, sin incluirla y hasta el valor especificado como Máximo de categorías para guardar. El nombre de raíz por defecto es *CI* y los nombres de campos por defecto son *CI_Lower_1*, *CI_Upper_1*, *CI_Lower_2*, *CI_Upper_2*, etcétera, que se corresponden con el orden de las categorías de destino.

**Residuos de Pearson** Guarda el residuo de Pearson de cada registro, que se puede utilizar tras el cálculo como diagnósticos del ajuste del modelo. El nombre del campo por defecto es *ResiduoPearson*.

**Confianzas.** Guarda la confianza en el valor pronosticado del destino categórico. La confianza calculada se puede basar en las probabilidades del valor pronosticado (la probabilidad más alta pronosticada) o la diferencia entre la probabilidad más alta pronosticada y la segunda probabilidad más alta pronosticada. El nombre del campo por defecto es *Confianza*.

**Exportar modelo.**Escribe el modelo en un archivo *.zip* externo. Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo. Especifique un nombre de archivo exclusivo y válido. Si la especificación de archivo hace referencia a un archivo existente, se sobreescribirá el archivo.

# Vista del modelo

Este procedimiento crea un objeto de modelo en el visor. Al activar (pulsando dos veces) este objeto se obtiene una vista interactiva del modelo.

Por defecto, se muestra la vista del resumen del modelo. Para ver otra vista del modelo, selecciónela de las miniaturas de vistas.

## Resumen del modelo

La vista es una instantánea, un resumen de un vistazo del modelo y su ajuste.

**Tabla.** La tabla identifica el destino, distribución de probabilidades y la función de enlace especificada en la configuración del destino. Si el destino está definido por eventos y ensayos, la casilla se divide mostrando el campo de eventos y el campo de ensayos o el número fijo de ensayos. Además se muestran el criterio de información de Akaike para muestras finitas (AICC) y el criterio de información bayesiano (BIC).

- **Akaike corregido.** Una medida para seleccionar y comparar modelos mixtos basada en la -2 log verosimilitud (restringida). Los valores menores indican modelos mejores. El AICC "corrige" el AIC respecto a tamaños muestrales pequeños. A medida que aumenta el tamaño muestral, el AICC converge con el AIC.
- **Bayesiano.** Una medida para seleccionar y comparar modelos basada en la -2 log verosimilitud. Los valores menores indican modelos mejores. El BIC también penaliza los modelos sobreparametrizados, pero de manera más estricta que el AIC.

**Gráfico.** Si el destino es categórico, un gráfico muestra la precisión del modelo final, que es el porcentaje de clasificaciones correctas.

## *Estructura de datos*

Proporciona un resumen de la estructura de datos especificada y le ayuda a comprobar que los sujetos y las medidas repetidas se han especificado correctamente. La información observada del primer sujeto se muestra para cada campo del destino y de las medidas repetidas, y el destino. Además, se muestra el número de niveles de cada campo de sujeto y el campo de medidas repetidas.

## *Pronóstico por observado*

En los destinos continuos, incluyendo destinos especificados como eventos/ensayos, muestra un diagrama de dispersión en intervalos de los valores pronosticados en el eje vertical por los valores observados en el eje horizontal. En teoría, los puntos deberían encontrarse en una línea de 45 grados; esta vista puede indicar si hay registros para los que el modelo realiza un pronóstico particularmente malo.

## *Clasificación*

En objetivos categóricos, muestra la clasificación cruzada de los valores observados frente a pronosticados en un mapa de calor y el porcentaje global correcto.

**Estilos de tabla.** Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable Estilo.

- **Porcentajes de fila.** Muestra los porcentajes de fila (los recuentos de casillas expresados como un porcentaje de los totales de filas) en las casillas. Ésta es la opción por defecto.
- **Recuentos de las casillas.** Muestra los recuentos de las casillas. El sombreado del mapa de calor se basa en los porcentajes de fila.
- **Mapa de calor.** No muestra los valores en las casillas, únicamente el sombreado.
- **Comprimido.** No muestra encabezados de fila y columna ni valores de las casillas. Puede ser útil si el objeto tiene múltiples categorías.

**Perdidos.** Si algunos de los registros tienen valores perdidos en el destino, se muestran en una fila (Perdidos) en todas las filas válidas. Los registros con valores perdidos no contribuyen en el porcentaje global correcto.

**Múltiples variables.** Si hay múltiples objetivos categóricos, cada destino se muestra en una tabla diferente y hay una lista desplegable Destino que controla los destinos que se muestran.

**Tablas grandes.** Si el destino que se muestra tiene más de 100 categorías, no se mostrará la tabla.

## *Efectos fijos*

Esta vista muestra el tamaño de cada efecto fijo en el modelo.

**Estilos.** Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable Estilo.

- **Diagrama.** Es un gráfico cuyos efectos se clasifican de superior a inferior en el orden en el que se han especificado en la configuración de Efectos fijos. Las líneas de conexión del diagrama se ponderan tomando como base la significación del efecto, con un grosor de línea mayor correspondiente a efectos con mayor significación (valores *p* inferiores). Ésta es la opción por defecto.
- **Tabla.** Se trata de una tabla ANOVA para el modelo completo y los efectos de modelo individuales. Los efectos individuales se clasifican de superior a inferior en el orden en el que se han especificado en la configuración de Efectos fijos.

**Significación.** Existe un control deslizante Significación que controla qué efectos se muestran en la vista. Se ocultan los efectos con valores de significación superiores al valor del control deslizante. Esto no cambia el modelo, simplemente le permite centrarse en los efectos más importantes. El valor por defecto es 1.00, de modo que no se filtran efectos tomando como base la significación.

## *Coeficientes fijos*

Esta vista muestra el valor de cada coeficiente fijo en el modelo. Tenga en cuenta que los factores (predictores categóricos) tienen codificación de indicador dentro del modelo, de modo que los **efectos** que contienen los factores generalmente tendrán múltiples **coeficientes** asociados: uno por cada categoría exceptuando la categoría que corresponde al coeficiente redundante.

**Estilos.** Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable Estilo.

- **Diagrama.** Es un gráfico que muestra la intersección primero y clasifica los efectos de superior a inferior en el orden en el que se han especificado en la configuración de Efectos fijos. Dentro de los efectos que contienen factores, los coeficientes se clasifican en orden ascendente de valores de datos. Las líneas de conexión del diagrama se colorean y se ponderan tomando como base la significación del coeficiente, con un grosor de línea mayor correspondiente a coeficientes con mayor significación (valores *p* inferiores). Este es el estilo por defecto.
- **Tabla.** Muestra los valores, las pruebas de significación y los intervalos de confianza para los coeficientes de modelos individuales. Tras la intersección, los efectos individuales se clasifican de superior a inferior en el orden en el que se han especificado en la configuración de Efectos fijos. Dentro de los efectos que contienen factores, los coeficientes se clasifican en orden ascendente de valores de datos.

**Multinomial.** Si la distribución multinomial está activada, la lista desplegable Multinomial controla los destinos categóricos que se mostrarán. El orden de clasificación de los valores de la lista está determinado por la especificación de la configuración de Opciones de construcción.

**Exponencial.** Muestra los cálculos del coeficiente exponencial y los intervalos de confianza de algunos tipos de modelos, incluyendo la regresión logística binaria (distribución binomial y enlace Logit), regresión logística nominal (distribución multinomial y enlace Logit), regresión binomial negativa (distribución binomial negativa y enlace Logit) y modelo Log-linear (distribución Poisson y enlace log).

**Significación.** Existe un control deslizante Significación que controla qué coeficientes se muestran en la vista. Se ocultan los coeficientes con valores de significación superiores al valor del control deslizante. Esto no cambia el modelo, simplemente le permite centrarse en los coeficientes más importantes. El valor por defecto es 1.00, de modo que no se filtran coeficientes tomando como base la significación.

## Covarianzas de efectos aleatorios

Esta vista muestra la matriz de covarianzas de efectos aleatorios (**G**).

**Estilos.** Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable Estilo.

- **Valores de covarianza.** Es un mapa de calor de la matriz de covarianza cuyos efectos se clasifican en el orden descendente en que se han especificado en la configuración de Efectos fijos. Los colores de corrgram se corresponden con los valores de la casilla tal y como se muestran en la clave. Ésta es la opción por defecto.
- **Corrgram.** Es un mapa de calor de la matriz de varianza.
- **Comprimido.** Es un mapa de calor de la matriz de covarianza sin los encabezados de fila y columna.

**Bloques.** Si hay múltiples de bloques de efectos aleatorios, hay una lista desplegable Bloque para seleccionar el bloque que se mostrará.

**Grupos.** Si un bloque de efecto aleatorio tiene una especificación de grupo, se incluirá una lista desplegable Grupo para seleccionar el nivel de grupo que se mostrará.

**Multinomial.** Si la distribución multinomial está activada, la lista desplegable Multinomial controla los destinos categóricos que se mostrarán. El orden de clasificación de los valores de la lista está determinado por la especificación de la configuración de Opciones de construcción.

## Parámetros de covarianza

Esta vista muestra los cálculos de los parámetros de covarianza y sus estadísticos relacionados de efectos residuales y aleatorios. Son resultados avanzados y fundamentales que proporcionan información sobre si la estructura de la covarianza es la adecuada.

**Tabla de resumen** Es una referencia rápida del número de parámetros en las matrices de covarianza residuales (**R**) y de efectos aleatorios (**G**), el rango (número de columnas) en el efecto fijo (**X**) y aleatorio (**Z**) matrices de diseño y el número de sujetos definidos por los campos de sujeto que definen la estructura de los datos.

**Tabla de parámetros de covarianza.** Para el efecto seleccionado, la estimación, error típico y el intervalo de confianza se muestran para cada parámetro de covarianza. El número de parámetros mostrados depende de la estructura de la covarianza del efecto y, para los bloques de efectos aleatorios, el número de efectos del bloque. Si ve que los parámetros no diagonales no son significativos, puede utilizar una estructura de covarianza más simple.

**Efectos** Si hay múltiples de bloques de efectos aleatorios, hay una lista desplegable Efecto para seleccionar el efecto de bloque residual o aleatorio que se mostrará. El efecto residual está siempre disponible.

**Grupos.** Si un bloque de efecto residual o aleatorio tiene una especificación de grupo, se incluirá una lista desplegable Grupo para seleccionar el nivel de grupo que se mostrará.

**Multinomial.** Si la distribución multinomial está activada, la lista desplegable Multinomial controla los destinos categóricos que se mostrarán. El orden de clasificación de los valores de la lista está determinado por la especificación de la configuración de Opciones de construcción.

## *Medias estimadas: Efectos significativos*

Son gráficos que muestran los 10 efectos fijos de todos los factores "más significativos", comenzando por interacciones de tres vías, a continuación interacciones de dos vías y, finalmente, efectos principales. El gráfico muestra el valor de estimación del modelo del destino en el eje vertical de cada valor del efecto principal (o primer efecto de la lista en una interacción) en el eje horizontal; se produce una línea diferente para cada valor del segundo efecto en una interacción y un gráfico distinto para cada valor del tercer efecto en una interacción de 3 vías; el resto de predictores se mantienen constantes. Proporciona una visualización útil de los efectos de los coeficientes de cada predictor en el objetivo. Tenga en cuenta que si no hay predictores significativos, no se generan medias estimadas.

**Confianza.** Muestra los límites superiores e inferiores de confianza de las medias marginales, utilizando el nivel de confianza especificado como parte de las Opciones de construcción.

## *Medias estimadas: Efectos personalizados*

Son tablas y gráficos de todos los efectos fijos y solicitados por el usuario.

**Estilos.** Existen varios estilos de visualización diferentes, que son accesibles desde la lista desplegable Estilo.

- **Diagrama.** Muestra un gráfico de líneas del valor de estimación del modelo del destino en el eje vertical de cada valor del efecto principal (o primer efecto de la lista en una interacción) en el eje horizontal; se produce una línea diferente para cada valor del segundo efecto en una interacción y un gráfico distinto para cada valor del tercer efecto en una interacción de 3 vías; el resto de predictores se mantienen constantes.

  Si se han solicitado contrastes, se muestra otro gráfico para comparar los niveles del campo de contraste; para las interacciones, se muestra un gráfico para cada nivel de combinación de los efectos diferente al campo de contraste. En contrastes **por parejas**, es una representación gráfica de la tabla de comparaciones en la que las distancias entre nodos de la red corresponden a las diferencias entre las muestras. Las líneas amarillas corresponden a diferencias estadísticamente importantes, mientras que las líneas negras corresponden a diferencias no significativas. Al pasar el ratón por una línea de la red se muestra una sugerencia con la significación corregida de la diferencia entre los nodos conectados por la línea.

En contrastes de **desviación**, se muestra un gráfico de barras con el valor estimado del modelo del destino en el eje vertical y los valores del campo de contraste en el eje horizontal; en las interacciones, se muestra un gráfico para cada combinación de niveles de los efectos en lugar del campo de contraste. Las barras muestran la diferencia entre cada nivel del campo de contraste y la media global, que se representa por una línea horizontal negra.

En contrastes **simples**, se muestra un gráfico de barras con el valor estimado del modelo del destino en el eje vertical y los valores del campo de contraste en el eje horizontal; en las interacciones, se muestra un gráfico para cada combinación de niveles de los efectos en lugar del campo de contraste. Las barras muestran la diferencia entre cada nivel del campo de contraste (excepto el último) y el último nivel, que se representa por una línea horizontal negra.

- **Tabla.** Este estilo muestra una tabla de valores del destino estimados por el usuario, su error típico y el intervalo de confianza de cada nivel de combinación de los campos en el efecto; el resto de predictores se mantienen constantes.

  Si se ha solicitado los contrastes, se muestra otra tabla con la estimación, error típico, pruebas de significancia e intervalo de confianza de cada contraste; en interacciones hay una lista diferente de filas para cada combinación de nivel de efectos en lugar del campo de contrastes. Además, se muestra una tabla con los resultados de las pruebas globales; en interacciones, hay una prueba global diferente para cada combinación de nivel de los efectos en lugar del campo de contraste.

**Confianza.** Cambia la visualización de los límites superiores e inferiores de confianza de las medias marginales, utilizando el nivel de confianza especificado como parte de las Opciones de construcción.

**Diseño.** Cambia la visualización del diagrama de contraste de parejas. El diseño circular revela menos contrastes que el diseño de red, pero evita que las líneas se superpongan.

Capítulo
9

# Análisis loglineal: Selección de modelo

El procedimiento de análisis loglineal de selección de modelo analiza tablas de contingencia de varios factores. Ajusta modelos loglineales jerárquicos a las tablas de contingencia multidimensionales utilizando un algoritmo de ajuste proporcional. Este procedimiento ayuda a encontrar cuáles de las variables categóricas están asociadas. Para construir los modelos se encuentran disponibles métodos de entrada forzada y de eliminación hacia atrás. Para los modelos saturados, es posible solicitar estimaciones de los parámetros y pruebas de asociación parcial. Un modelo saturado añade 0,5 a todas las casillas.

**Ejemplo.** En un estudio sobre las preferencias del consumidor por uno de entre dos detergentes, los investigadores contaron las personas presentes en cada grupo, combinando las diversas categorías de grado de dureza del agua (blanda, media o dura), uso previo de una de las dos marcas y temperaturas de lavado (frío o caliente). Averiguaron que la temperatura está relacionada con la dureza del agua y con la preferencia por una u otra marca.

**Estadísticos.** Frecuencias, residuos, estimaciones de los parámetros, errores típicos, intervalos de confianza y pruebas de asociación parcial. Para los modelos personalizados, gráficos de residuos y gráficos de probabilidad normal.

**Datos.** Las variables de factor son categóricas. Todas las variables que se vayan a analizar deben ser numéricas. Las variables categóricas de cadena se pueden recodificar en variables numéricas antes de comenzar el análisis para la selección del modelo.

Evite especificar muchas variables con un número elevado de niveles. Tales especificaciones pueden conducir a una situación en la que muchas casillas posean un número reducido de observaciones y los valores de chi-cuadrado puede que no sean útiles.

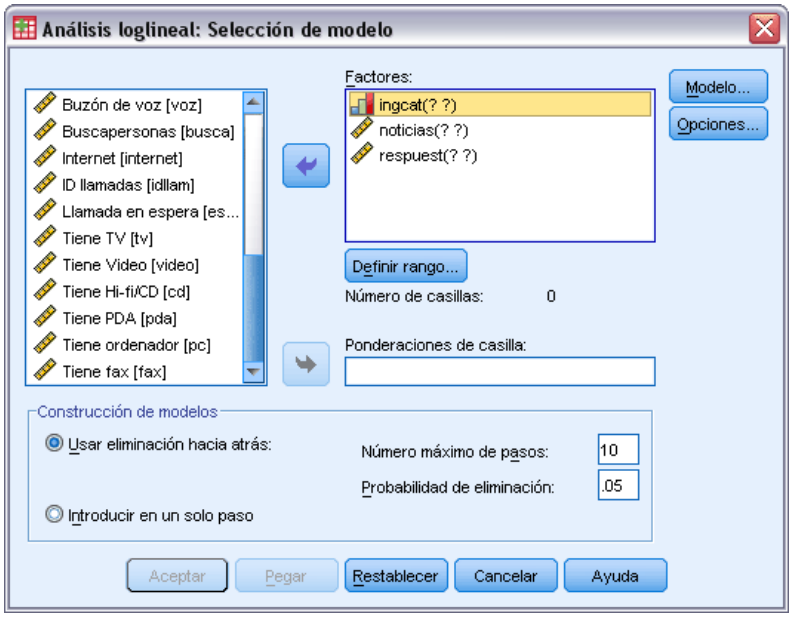
**Procedimientos relacionados.** El procedimiento Selección de modelo puede ayudar a identificar los términos que se necesitan en el modelo. A continuación, puede pasar a evaluar el modelo utilizando el Análisis loglineal general o el Análisis loglineal logit. Es posible utilizar la recodificación automática para recodificar las variables de cadena. Si una variable numérica posee categorías vacías, utilice Recodificar para crear valores enteros consecutivos.

### *Para obtener una selección de modelo en el análisis loglineal*

En los menús, seleccione:

Analizar > Loglineal > Selección de modelo...

Figura 9-1
*Cuadro de diálogo Análisis loglineal: Selección de modelo*

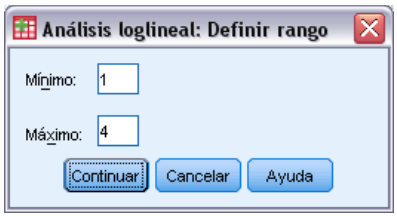



- Seleccione dos o varios factores categóricos numéricos.
- Seleccione una o más variables de factor en la lista Factores y pulse en Definir rango.
- Defina el rango de valores para cada variable de factor.
- Seleccione una opción en la sección Construcción de modelos.

  Si lo desea, puede seleccionar una variable de ponderación de casilla para especificar los ceros estructurales.

## *Análisis loglineal: Definir rango*

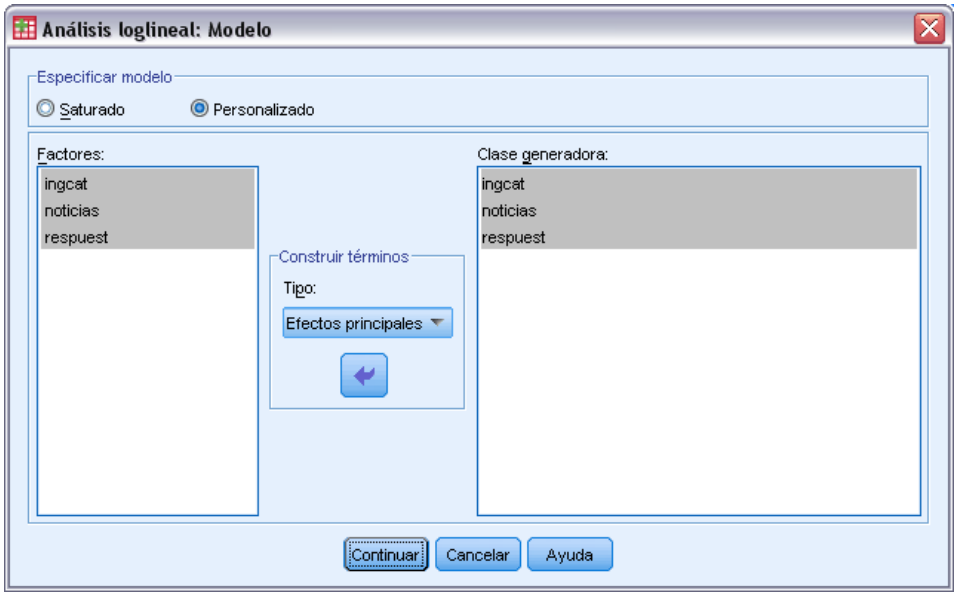
Figura 9-2
*Cuadro de diálogo Análisis loglineal: Definir rango*




Se debe indicar el rango de categorías para cada variable de factor. Los valores para Mínimo y Máximo corresponden a las categorías menor y mayor de la variable de factor. Ambos valores deben ser enteros y el valor mínimo debe ser menor que el máximo. Se excluyen los casos con valores fuera de los límites. Por ejemplo, si especifica un valor mínimo de 1 y uno máximo de 3, solamente se utilizarán los valores 1, 2 y 3. Repita este proceso para cada variable de factor.

# Análisis loglineal: Modelo

Figura 9-3
*Cuadro de diálogo Análisis loglineal: Modelo*




**Especificar modelo.** Un modelo saturado contiene todos los efectos principales de factor y todas las interacciones factor por factor. Seleccione Personalizado para especificar una clase generadora para un modelo no saturado.

**Clase generadora.** Una clase generadora es una lista de los términos de mayor orden en los que se encuentran implicados los factores. Un modelo jerárquico contiene los términos que definen la clase generadora y todos los relativos de orden inferior. Supongamos que se seleccionan las variables *A*, *B* y *C* en la lista Factores y, a continuación, Interacción en la lista desplegable Construir términos. El modelo resultante contendrá la interacción triple *A*B*C* especificada, las interacciones dobles *A*B*, *A*C* y *B*C*, así como los efectos principales para *A*, *B* y *C*. No especifique los relativos de orden inferior en la clase generadora.

## Construir términos

Para las covariables y los factores seleccionados:

**Interacción.** Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el método por defecto.

**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.

# Análisis loglineal: Opciones

Figura 9-4
*Cuadro de diálogo Análisis loglineal: Opciones*


**Mostrar.** Puede elegir entre Frecuencias, Residuos, o ambos. En un modelo saturado, las frecuencias observadas y las esperadas son iguales, y los residuos son iguales a 0.

**Gráfico.** Para los modelos personalizados es posible elegir uno o ambos tipos de gráficos, Residuos y Probabilidad normal. Éstos ayudarán a determinar cómo se ajusta el modelo a los datos.

**Mostrar para el modelo saturado.** Para un modelo saturado, es posible elegir Estimaciones de los parámetros. Las estimaciones de los parámetros pueden ayudar a determinar qué términos se pueden excluir del modelo. También se encuentra disponible una tabla de asociación que enumera pruebas de asociación parcial. Esta opción supone un proceso de cálculo muy extenso cuando se trata de tablas con muchos factores.

**Criterios del modelo.** Se utiliza un algoritmo iterativo de ajuste proporcional para obtener las estimaciones de los parámetros. Es posible suprimir uno o más criterios de estimación especificando N° máximo de iteraciones, Convergencia o Delta (un valor añadido a todas las frecuencias de casilla para los modelos saturados).

# Funciones adicionales del comando HILOGLINEAR

Con el lenguaje de sintaxis de comandos también podrá:

- Especificar las ponderaciones de casilla en forma de matriz (utilizando el subcomando `CWEIGHT`).
- Generar análisis de varios modelos con un único comando (utilizando el subcomando `DESIGN`).

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

Capítulo
10

# Análisis loglineal general

El procedimiento Análisis loglineal general analiza las frecuencias de las observaciones incluidas en cada categoría de la clasificación cruzada de una tabla de contingencia. Cada una de las clasificaciones cruzadas de la tabla constituye una casilla y cada variable categórica se denomina factor. La variable dependiente es el número de casos (la frecuencia) en una casilla de la tabla de contingencia y las variables explicativas son los factores y las covariables. Este procedimiento estima los parámetros de máxima verosimilitud de modelos loglineales jerárquicos y no jerárquicos utilizando el método de Newton-Raphson. Es posible analizar una distribución multinomial o de Poisson.

Se pueden seleccionar hasta 10 factores para definir las casillas de una tabla. Una variable de estructura de casilla permite definir ceros estructurales para tablas incompletas, incluir en el modelo un término de desplazamiento, ajustar un modelo log-tasa o implementar el método de corrección de las tablas marginales. Las variables de contraste permiten el cálculo del logaritmo de la razón de ventajas generalizadas (GLOR).

Se muestra automáticamente información sobre el modelo y estadísticos de bondad de ajuste. Además es posible mostrar una variedad de estadísticos y gráficos, o guardar los valores pronosticados y los residuos en el conjunto de datos activo.

**Ejemplo.** Los datos de un informe sobre accidentes de automóviles en Florida se utilizan para determinar la relación existente entre el hecho de llevar puesto el cinturón de seguridad y si el daño fue mortal o no. La razón de las ventajas indica la evidencia significativa de una relación.

**Estadísticos.** Frecuencias esperadas y observadas; residuos de desviación, corregidos y brutos; matriz del diseño; estimaciones de los parámetros; razón de las ventajas; log-razón de las ventajas; GLOR (log-razón de las ventajas generalizada); estadístico de Wald; intervalos de confianza. Gráficos: residuos corregidos, residuos de desviación y probabilidad normal.

**Datos.** Los factores son categóricos y las covariables de casilla son continuas. Cuando se introduce una covariable en el modelo, se aplica a cada casilla el valor medio de la covariable para los casos de esa casilla. Las variables de contraste son continuas. Se utilizan para calcular los logaritmos de la razón de las ventajas generalizadas. Los valores de la variable de contraste son los coeficientes para la combinación lineal de los logaritmos de las frecuencias esperadas de casilla.

Una variable de estructura de casilla asigna ponderaciones. Por ejemplo, si algunas de las casillas son ceros estructurales, la variable de estructura de casilla posee un valor de 0 ó 1. No utilice una variable de estructura de casilla para ponderar los datos agregados. En su lugar, elija Ponderar casos en el menú Datos.

**Supuestos.** Existen dos distribuciones disponibles en el análisis loglineal general: Poisson y multinomial.

Bajo el supuesto de distribución de Poisson:

- El tamaño total de la muestra no se fija antes del estudio o el análisis no es condicional al tamaño total de la muestra.
- El evento de una observación que está en una casilla es estadísticamente independiente de los recuentos de casilla de otras casillas.

Bajo el supuesto de distribución multinomial:

- El tamaño muestral total es fijo o el análisis está condicionado al tamaño muestral total.
- Los recuentos de casilla no son estadísticamente independientes.

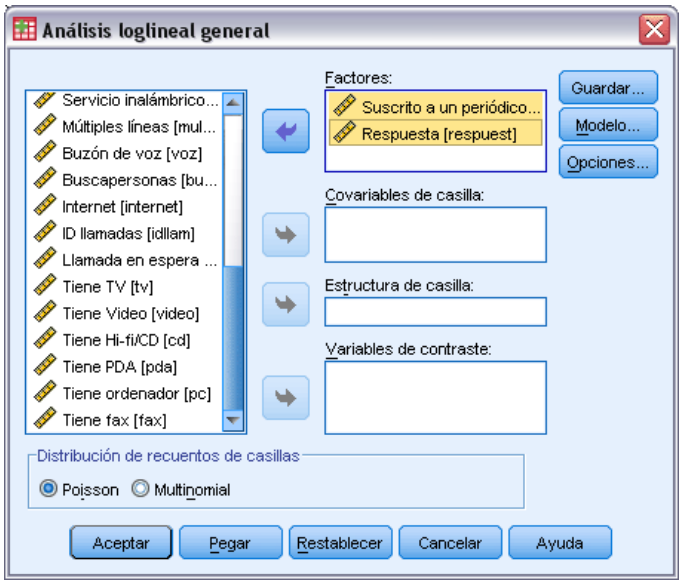
**Procedimientos relacionados.** Utilice el procedimiento Tablas de contingencia para examinar las tablas de contingencia. Emplee el procedimiento Loglineal logit cuando resulte natural considerar una o más variables categóricas como variables de respuesta y las demás como variables explicativas.

### *Para obtener un análisis loglineal general*

▶ En los menús, seleccione:

Analizar > Loglineal > General...

Figura 10-1
*Cuadro de diálogo Análisis loglineal general*



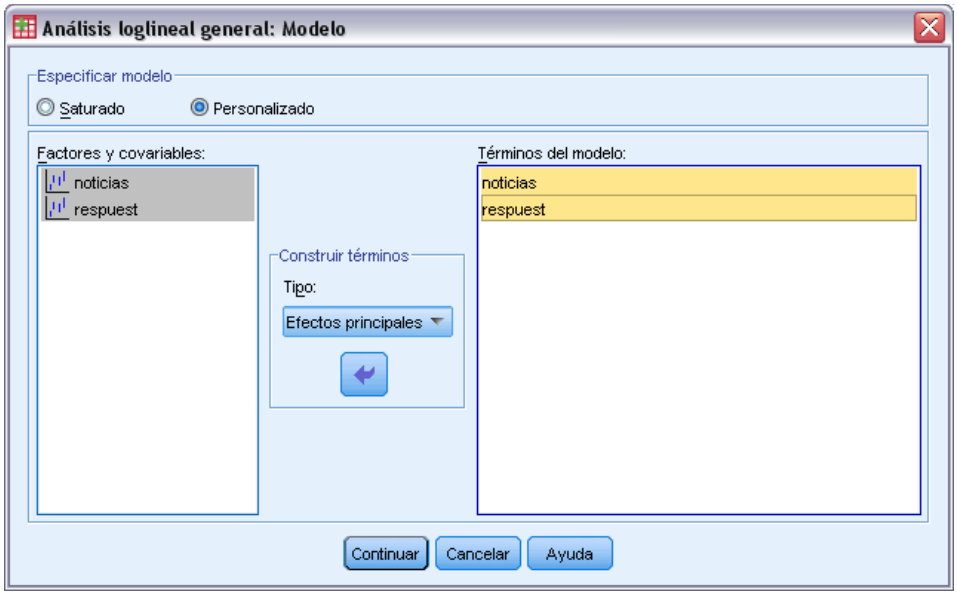

▶ En el cuadro de diálogo Análisis loglineal general, seleccione un máximo de diez variables de factor.

Si lo desea, puede:

- Seleccionar covariables de casilla.
- Seleccionar una variable de estructura de casilla para definir ceros estructurales o incluir un término de desplazamiento.
- Seleccionar una variable de contraste.

# Análisis loglineal general: Modelo

Figura 10-2
*Cuadro de diálogo Análisis loglineal general: Modelo*




**Especificar modelo.** Un modelo saturado contiene todos los efectos principales e interacciones que impliquen a las variables de factor. No contiene términos para las covariables. Seleccione Personalizado para especificar sólo un subconjunto de interacciones o para especificar interacciones factor por covariable.

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

**Términos del modelo.** El modelo depende de la naturaleza de los datos. Después de seleccionar Personalizado, puede elegir los efectos principales y las interacciones que sean de interés para el análisis. Indique todos los términos que desee incluir en el modelo.

## Construir términos

Para las covariables y los factores seleccionados:

**Interacción.** Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el método por defecto.

**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.

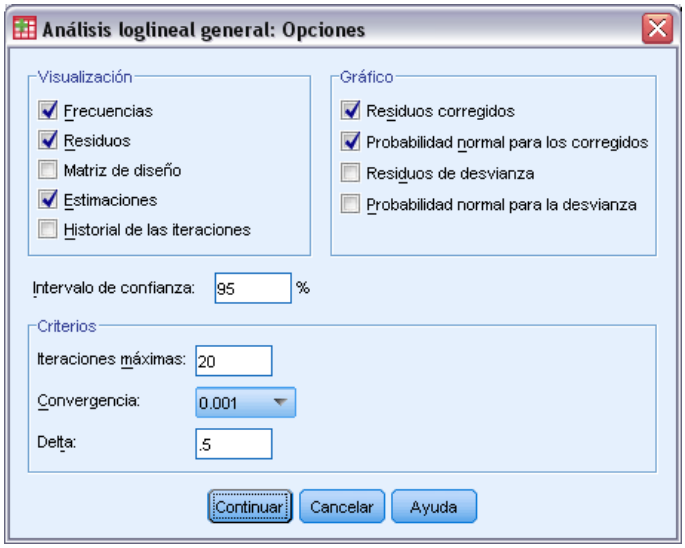
**Todas de 3.** Crea todas las interacciones triples posibles de las variables seleccionadas.

**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.

# Análisis loglineal general: Opciones

Figura 10-3
*Cuadro de diálogo Análisis loglineal general: Opciones*




El procedimiento Análisis loglineal general muestra información sobre el modelo y los estadísticos de bondad de ajuste. Además, tiene la posibilidad de elegir una o varias de las opciones siguientes:

**Mostrar.** Puede elegir entre varias opciones de estadísticos: frecuencias esperadas y observadas de casilla, residuos de desviación, corregidos y simples (o brutos), una matriz del diseño del modelo y estimaciones de los parámetros para el modelo.

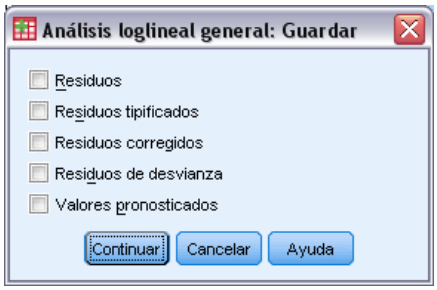
**Gráfico.** Los gráficos, los cuales sólo están disponibles para los modelos personalizados, incluyen dos diagramas de dispersión matriciales: residuos corregidos o residuos de desviación respecto a los recuentos observados y los esperados de las casillas. Además es posible mostrar gráficos de probabilidad normal y gráficos normales sin tendencia de los residuos de desviación o corregidos.

**Intervalo de confianza.** Se puede ajustar el intervalo de confianza para las estimaciones de los parámetros.

**Criterios.** Se utiliza el método de Newton-Raphson para obtener estimaciones maximo-verosímiles de los parámetros. Es posible introducir nuevos valores para el número máximo de iteraciones, el criterio de convergencia y la delta (constante añadida a todas las casillas para las aproximaciones iniciales). La delta permanece en las casillas para los modelos saturados.

# Análisis loglineal general: Guardar

Figura 10-4
*Cuadro de diálogo Análisis loglineal general: Guardar*




Seleccione los valores que desee guardar como nuevas variables en el conjunto de datos activo. El sufijo *n* añadido a los nuevos nombres de variable se incrementa para formar un nombre exclusivo para cada variable guardada.

Los valores guardados hacen referencia a los datos agregados (las casillas de la tabla de contingencia), aunque los datos estén registrados como observaciones individuales en el Editor de datos. Si se guardan los valores pronosticados o los residuos para datos no agregados, el valor a guardar para una casilla de la tabla de contingencia es introducido en el Editor de datos para cada caso de esa casilla. Para que los valores guardados tengan sentido, se debería agregar los datos para obtener los recuentos de casilla.

Se pueden guardar cuatro tipos de residuos: de desviación, corregidos, tipificados y brutos. También se pueden guardar los valores pronosticados.

- **Residuos.** También llamado residuo simple o bruto, es la diferencia entre la frecuencia observada para la casilla y su frecuencia esperada.
- **Residuos tipificados.** Los residuos divididos por una estimación de su error típico. Los residuos tipificados se conocen también como residuos de Pearson.
- **Residuos corregidos.** El residuo tipificado dividido por la estimación de su error típico. Dado que, cuando el modelo es el correcto, los residuos corregidos son asintóticamente normales típicos, éstos son preferidos a los residuos tipificados a la hora de contrastar la normalidad.
- **Residuos de desvianza.** Raíz cuadrada, con signo, de la contribución individual al estadístico de chi-cuadrado de la razón de verosimilitud (G al cuadrado), donde el signo es el signo del residuo (frecuencia observada menos frecuencia esperada). Los residuos de desviación tienen una distribución normal típica asintótica.

# Funciones adicionales del comando GENLOG

Con el lenguaje de sintaxis de comandos también podrá:

- Calcular combinaciones lineales de las frecuencias observadas de casilla y las frecuencias esperadas de casilla e imprimir los residuos, residuos tipificados y residuos corregidos de esa combinación (utilizando el subcomando `GERESID`).
- Cambiar el valor por defecto del umbral para la comprobación de la redundancia (utilizando el subcomando `CRITERIA`).
- Mostrar los residuos tipificados (utilizando el subcomando `PRINT`).

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

Capítulo
11

# Análisis loglineal logit

El procedimiento Análisis loglineal logit analiza la relación entre variables dependientes (o de respuesta) y variables independientes (o explicativas). Las variables dependientes siempre son categóricas, mientras que las variables independientes pueden ser categóricas (factores). Otras variables independientes, las covariables de casilla, pueden ser continuas pero no se aplican en forma de caso por caso. A una casilla dada se le aplica la media ponderada de la covariable para los casos de esa casilla. El logaritmo de las ventajas de las variables dependientes se expresa como una combinación lineal de parámetros. Se supone automáticamente una distribución multinomial; estos modelos se denominan a veces modelos logit multinomiales. Este procedimiento estima los parámetros de los modelos loglineales logit utilizando el algoritmo de Newton-Raphson.

Es posible seleccionar de 1 a 10 variables dependientes y de factor en combinación. Una variable de estructura de casilla permite definir ceros estructurales para tablas incompletas, incluir en el modelo un término de desplazamiento, ajustar un modelo log-tasa o implementar el método de corrección de las tablas marginales. Las variables de contraste permiten el cálculo del logaritmo de la razón de ventajas generalizadas (GLOR). Los valores de la variable de contraste son los coeficientes para la combinación lineal de los logaritmos de las frecuencias esperadas de casilla.

Se muestra automáticamente información sobre el modelo y estadísticos de bondad de ajuste. Además es posible mostrar una variedad de estadísticos y gráficos, o guardar los valores pronosticados y los residuos en el conjunto de datos activo.

**Ejemplo.** En un estudio en Florida se incluyeron 219 caimanes. ¿Cómo varía el tipo de comida de los caimanes en función del tamaño del caimán y de los cuatro lagos en los que viven? Los resultados del estudio mostraron que la ventaja para un caimán pequeño preferir reptiles a peces es 0,70 veces menor que para un caimán grande; además la ventaja de preferir fundamentalmente reptiles en vez de peces fue más alta en el lago 3.

**Estadísticos.** Frecuencias observadas y esperadas; residuos brutos, corregidos y de desviación; matriz de diseño; estimaciones de los parámetros; logaritmo de la razón de las ventajas generalizadas; estadístico de Wald; intervalos de confianza. Gráficos: residuos corregidos, residuos de desviación y gráficos de probabilidad normal.

**Datos.** Las variables dependientes son categóricas. Los factores son categóricos; pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Las covariables de casilla pueden ser continuas, pero cuando una covariable está en el modelo, se aplica a una casilla dada el valor medio de la covariable para los casos de a esa casilla. Las variables de contraste son continuas. Se utilizan para calcular el logaritmo de la razón de las ventajas (GLOR). Los valores de la variable de contraste son los coeficientes para la combinación lineal de los logaritmos de las frecuencias esperadas de casilla.

Una variable de estructura de casilla asigna ponderaciones. Por ejemplo, si algunas de las casillas son ceros estructurales, la variable de estructura de casilla posee un valor de 0 o 1. No utilice una variable de estructura de casilla para ponderar datos de agregación. En su lugar, utilice Ponderar casos del menú Datos.

**Supuestos.** Se supone que los recuentos dentro de cada combinación de categorías de las variables explicativas poseen una distribución multinomial. Bajo el supuesto de distribución multinomial:

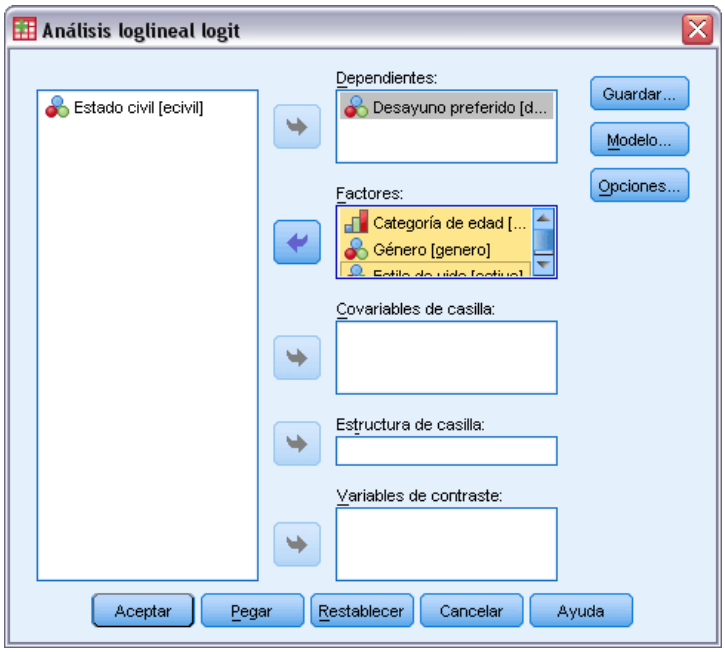
- El tamaño muestral total es fijo o el análisis está condicionado al tamaño muestral total.
- Los recuentos de casilla no son estadísticamente independientes.

**Procedimientos relacionados.** Utilice el procedimiento Tablas de contingencia para mostrar las tablas de contingencia. Utilice el procedimiento Análisis loglineal general cuando quiera analizar la relación entre una frecuencia observada y un conjunto de variables explicativas.

### *Para obtener un análisis loglineal logit*

▶ En los menús, seleccione:
Analizar > Loglineal > Logit...

Figura 11-1
*Cuadro de diálogo Análisis loglineal logit*




▶ En el cuadro de diálogo Análisis loglineal logit, seleccione una o más variables dependientes.

▶ Seleccione una o más variables de factor.

El número total de variables dependientes y de factor debe ser menor o igual a 10.

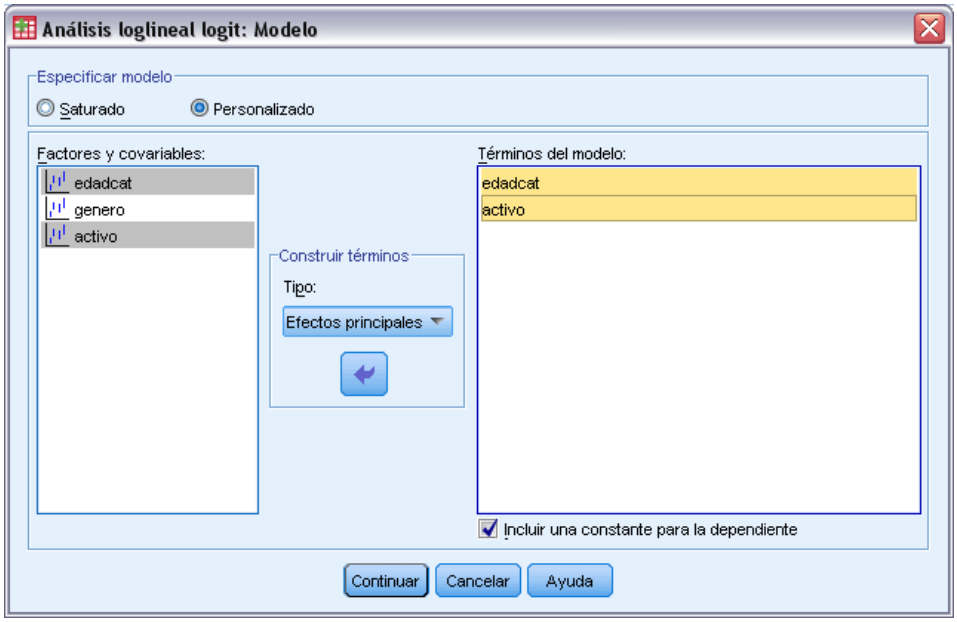
Si lo desea, puede:

- Seleccionar covariables de casilla.

- Seleccionar una variable de estructura de casilla para definir ceros estructurales o incluir un término de desplazamiento.
- Seleccionar una o más variables de contraste.

## *Análisis loglineal logit: Modelo*

Figura 11-2
*Cuadro de diálogo Análisis loglineal logit: Modelo*




**Especificar modelo.** Un modelo saturado contiene todos los efectos principales e interacciones que impliquen a las variables de factor. No contiene términos para las covariables. Seleccione Personalizado para especificar sólo un subconjunto de interacciones o para especificar interacciones factor por covariable.

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

**Términos del modelo.** El modelo depende de la naturaleza de los datos. Después de seleccionar Personalizado, puede elegir los efectos principales y las interacciones que sean de interés para el análisis. Indique todos los términos que desee incluir en el modelo.

Los términos se añaden al diseño tomando todas las combinaciones posibles de los términos dependientes y haciendo emparejando cada combinación con cada término de la lista de términos del modelo. Si se selecciona la opción Incluir una constante para la dependiente, también se añade un término unidad (1) a la lista del modelo.

Por ejemplo, supongamos que las variables *D1* y *D2* son las variables dependientes. El procedimiento Análisis loglineal logit crea una lista de términos dependientes (*D1*, *D2*, *D1*D2*). Si la lista Términos del modelo contiene *M1* y *M2* y se incluye una constante, la lista del modelo contendrá 1, *M1* y *M2*. El diseño resultante incluye combinaciones de cada término del modelo con cada término dependiente:

*D1*, *D2*, *D1*D2*

*M1*D1*, *M1*D2*, *M1*D1*D2*

*M2*D1*, *M2*D2*, *M2*D1*D2*

**Incluir una constante para la dependiente.** Incluye una constante para la variable dependiente en un modelo personalizado.

### *Construir términos*

Para las covariables y los factores seleccionados:

**Interacción.** Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el método por defecto.

**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.

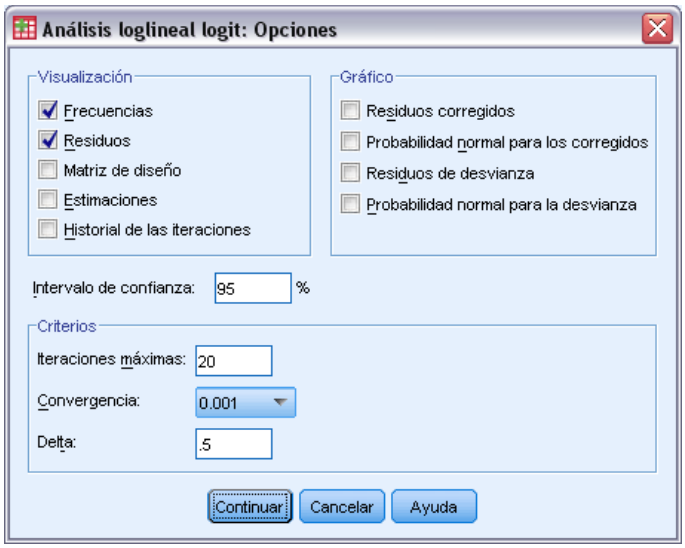
**Todas de 3.** Crea todas las interacciones triples posibles de las variables seleccionadas.

**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.

## *Análisis loglineal logit: Opciones*

Figura 11-3
*Cuadro de diálogo Análisis loglineal logit: Opciones*




El procedimiento Análisis loglineal logit muestra información sobre el modelo y estadísticos de bondad de ajuste. Además, es posible elegir una o más de las siguientes opciones:

**Mostrar.** Existen varios estadísticos disponibles para su presentación: Frecuencias de casillas observadas y esperadas; Residuos de desviación, corregidos y brutos; Matriz del diseño del modelo y Estimaciones de parámetros para el modelo.

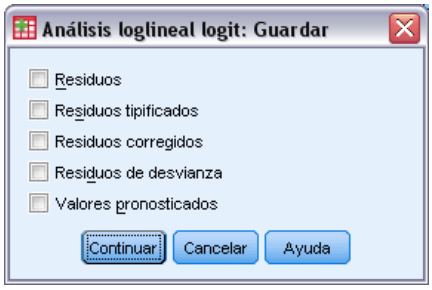
**Gráfico.** Los gráficos disponibles para los modelos personalizados incluyen dos diagramas de dispersión matriciales (los residuos corregidos o los residuos de desviación respecto a las frecuencias de casilla observadas y esperadas). Además es posible mostrar gráficos de probabilidad normal y gráficos normales sin tendencia de los residuos de desviación o corregidos.

**Intervalo de confianza.** Se puede ajustar el intervalo de confianza para las estimaciones de los parámetros.

**Criterios.** Se utiliza el método de Newton-Raphson para obtener estimaciones maximo-verosímiles de los parámetros. Es posible introducir nuevos valores para el número máximo de iteraciones, el criterio de convergencia y la delta (constante añadida a todas las casillas para las aproximaciones iniciales). La delta permanece en las casillas para los modelos saturados.

# *Análisis loglineal logit: Guardar*

Figura 11-4
*Cuadro de diálogo Análisis loglineal logit: Guardar*




Seleccione los valores que desee guardar como nuevas variables en el conjunto de datos activo. El sufijo *n* añadido a los nuevos nombres de variable se incrementa para formar un nombre exclusivo para cada variable guardada.

Los valores guardados hacen referencia a los datos agregados (a casillas de la tabla de contingencia), aunque los datos se encuentren registrados como observaciones individuales en el Editor de datos. Si se guardan los valores pronosticados o los residuos para datos no agregados, el valor a guardar para una casilla de la tabla de contingencia es introducido en el Editor de datos para cada caso de esa casilla. Para que los valores guardados tengan sentido, se debería agregar los datos para obtener los recuentos de casilla.

Se pueden guardar cuatro tipos de residuos: de desviación, corregidos, tipificados y brutos. También se pueden guardar los valores pronosticados.

- **Residuos.** También llamado residuo simple o bruto, es la diferencia entre la frecuencia observada para la casilla y su frecuencia esperada.
- **Residuos tipificados.** Los residuos divididos por una estimación de su error típico. Los residuos tipificados se conocen también como residuos de Pearson.

- **Residuos corregidos.** El residuo tipificado dividido por la estimación de su error típico. Dado que, cuando el modelo es el correcto, los residuos corregidos son asintóticamente normales típicos, éstos son preferidos a los residuos tipificados a la hora de contrastar la normalidad.
- **Residuos de desvianza.** Raíz cuadrada, con signo, de la contribución individual al estadístico de chi-cuadrado de la razón de verosimilitud (G al cuadrado), donde el signo es el signo del residuo (frecuencia observada menos frecuencia esperada). Los residuos de desviación tienen una distribución normal típica asintótica.

## Funciones adicionales del comando GENLOG

Con el lenguaje de sintaxis de comandos también podrá:

- Calcular combinaciones lineales de las frecuencias observadas de casilla y las frecuencias esperadas de casilla e imprimir los residuos, residuos tipificados y residuos corregidos de esa combinación (utilizando el subcomando `GERESID`).
- Cambiar el valor por defecto del umbral para la comprobación de la redundancia (utilizando el subcomando `CRITERIA`).
- Mostrar los residuos tipificados (utilizando el subcomando `PRINT`).

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

Capítulo
# 12

# Tablas de mortalidad

Existen muchas situaciones en las se desea examinar la distribución de un período entre dos eventos, como la duración del empleo (tiempo transcurrido entre el contrato y el abandono de la empresa). Sin embargo, este tipo de datos suele incluir algunos casos para los que no se registra el segundo evento; por ejemplo, la gente que todavía trabaja en la empresa al final del estudio. Las razones para que no se verifique el segundo evento pueden ser muy variadas: en algunos casos, el evento simplemente no tiene lugar antes de que finalice el estudio; en otros, el investigador puede haber perdido el seguimiento de su estado en algún momento anterior a que finalice el estudio; y existen además casos que no pueden continuar por razones ajenas al estudio (como el caso en que un empleado caiga enfermo y se acoja a una baja laboral). Estos casos se conocen globalmente como **casos censurados** y hacen que el uso de técnicas tradicionales como las pruebas *t* o la regresión lineal sea inapropiado para este tipo de estudio.

Existe una técnica estadística útil para este tipo de datos llamada **tabla de mortalidad** de "seguimiento". La idea básica de la tabla de mortalidad es subdividir el período de observación en intervalos de tiempo más pequeños. En cada intervalo, se utiliza toda la gente que se ha observado como mínimo durante ese período de tiempo para calcular la probabilidad de que un evento terminal tenga lugar dentro de ese intervalo. Las probabilidades estimadas para cada intervalo se utilizan para estimar la probabilidad global de que el evento tenga lugar en diferentes puntos temporales.

**Ejemplo.** ¿Funciona la nueva terapia de parches de nicotina mejor que la terapia de parches tradicional a la hora de ayudar a la gente a dejar de fumar? Se podría llevar a cabo un estudio utilizando dos grupos de fumadores, uno que haya seguido la terapia tradicional y el otro la terapia experimental. Al construir las tablas de mortalidad a partir de los datos podrá comparar las tasas de abstinencia globales para los dos grupos, con el fin de determinar si el tratamiento experimental representa una mejora con respecto a la terapia tradicional. Si desea obtener información más detallada, también es posible representar gráficamente las funciones de impacto o de supervivencia y compararlas visualmente.

**Estadísticos.** Número que entra, número que abandona, número expuesto a riesgo, número de eventos terminales, proporción que termina, proporción que sobrevive, proporción acumulada que sobrevive (y error típico), densidad de probabilidad (y error típico), tasa de impacto (y error típico) para cada intervalo de tiempo en cada grupo. Gráficos: gráficos de las funciones para supervivencia, log de la supervivencia, densidad, tasa de impacto y uno menos la supervivencia.

**Datos.** La variable de tiempo deberá ser cuantitativa. La variable de estado deberá ser dicotómica o categórica, codificada en forma de números enteros, con los eventos codificados en forma de un valor único o un rango de valores consecutivos. Las variables de factor deberán ser categóricas, codificadas como valores enteros.

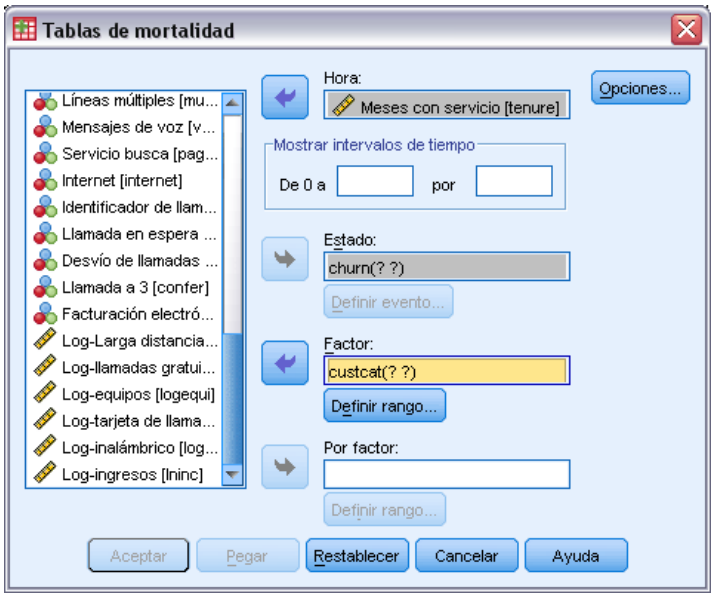
**Supuestos.** Las probabilidades para el evento de interés deben depender solamente del tiempo transcurrido desde el evento inicial (se asume que son estables con respecto al tiempo absoluto). Es decir, los casos que se introducen en el estudio en horas diferentes (por ejemplo, pacientes que inician el tratamiento en horas diferentes) se deberían comportar de manera similar. Tampoco deben existir diferencias sistemáticas entre los casos censurados y los no censurados. Si, por ejemplo, muchos de los casos censurados son pacientes en condiciones más graves, los resultados pueden resultar sesgados.

**Procedimientos relacionados.** El procedimiento Tablas de mortalidad utiliza un enfoque actuarial en esta clase de análisis (conocido de manera genérica como Análisis de supervivencia). El procedimiento Análisis de supervivencia de Kaplan-Meier utiliza un método ligeramente diferente para calcular las tablas de mortalidad, el cual no se basa en la partición del período de observación en intervalos de tiempo más pequeños. Este método es recomendable si se tiene un número pequeño de observaciones, de manera que habrá solamente un pequeño número de observaciones en cada intervalo de tiempo de supervivencia. Si dispone de variables que cree que están relacionadas con el tiempo de supervivencia o variables que desea controlar (covariables), utilice el procedimiento Regresión de Cox. Si las covariables pueden tener distintos valores en diferentes puntos temporales para el mismo caso, utilice el procedimiento Regresión de Cox con covariables dependientes del tiempo.

### *Para crear una tabla de mortalidad*

- En los menús, seleccione:
  Analizar > Supervivencia > Tablas de mortalidad...

Figura 12-1
*Cuadro de diálogo Tablas de mortalidad*




- Seleccione una variable **numérica** de supervivencia.
- Especifique los intervalos de tiempo que se van a examinar.

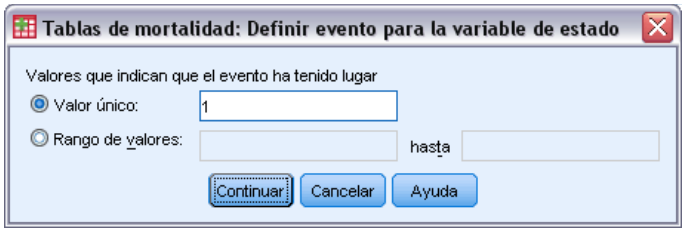
- Seleccione una variable de estado para definir casos para los que tuvo lugar el evento terminal.
- Pulse en Definir evento para especificar el valor de la variable de estado, el cual indica que ha tenido lugar un evento.

  Si lo desea, puede seleccionar una variable de factor de primer orden. Se generan tablas actuariales de la variable de supervivencia para cada categoría de la variable de factor.

  Además es posible seleccionar una variable *por factor* de segundo orden. Las tablas actuariales de la variable de supervivencia se generan para cada combinación de las variables de factor de primer y segundo orden.

## Tablas de mortalidad: Definir evento para la variable de estado

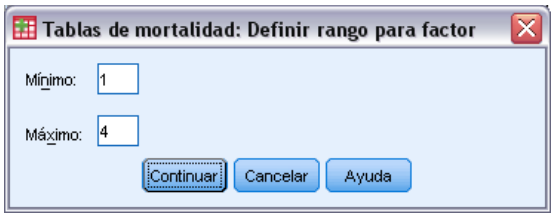
Figura 12-2
*Cuadro de diálogo Tablas de mortalidad: Definir evento para la variable de estado*




Las apariciones del valor o valores seleccionados para la variable de estado indican que el evento terminal ha tenido lugar para esos casos. Todos los demás casos se consideran censurados. Introduzca un único valor o un rango de valores que identifiquen el evento de interés.

## Tablas de mortalidad: Definir rango

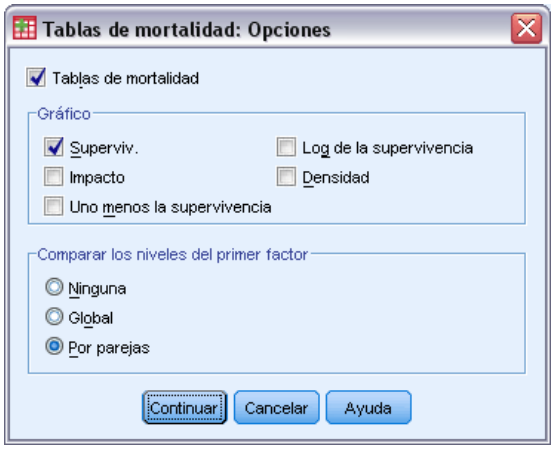
Figura 12-3
*Cuadro de diálogo Tablas de mortalidad: Definir rango*




Los casos con valores para la variable de factor dentro del rango especificado se incluirán en el análisis y se generarán tablas individuales (y gráficos si se solicita) para cada valor individual dentro del rango.

# Tablas de mortalidad: Opciones

Figura 12-4
*Cuadro de diálogo Tablas de mortalidad: Opciones*




Es posible controlar diversos aspectos del análisis de Tablas de mortalidad.

**Tablas de mortalidad.** Para suprimir la presentación de las tablas de mortalidad en los resultados, desactive Tablas de mortalidad.

**Gráfico.** Permite solicitar gráficos de las funciones de supervivencia. Si se han definido variables de factor, se generan gráficos para cada subgrupo definido por las variables de factor. Los gráficos disponibles son Supervivencia, Log de la supervivencia, Impacto, Densidad y Uno menos la supervivencia.

- **Supervivencia.** Muestra la función de supervivencia acumulada, en una escala lineal.
- **Log de la supervivencia.** Muestra la función de supervivencia, en una escala logarítmica.
- **Impacto.** Muestra la función de impacto acumulada, en una escala lineal.
- **Densidad.** Muestra la función de densidad.
- **Uno menos la supervivencia.** Representa la función uno menos la supervivencia en una escala lineal.

**Comparar los niveles del primer factor.** Si tiene una variable de control de primer orden, se puede seleccionar una de las opciones de este grupo para realizar la prueba de Wilcoxon (Gehan), la cual compara la supervivencia para los subgrupos. Las pruebas se realizan en el factor de primer orden. Si ha definido un factor de segundo orden, se realizarán pruebas para cada nivel de la variable de segundo orden.

# Funciones adicionales del comando SURVIVAL

Con el lenguaje de sintaxis de comandos también podrá:

- Especificar más de una variable dependiente.
- Especificar intervalos espaciados de forma desigual.
- Especificar más de una variable de estado.

- Especificar comparaciones que no incluyan todas las variables de control y de factor.
- Calcular comparaciones aproximadas, no exactas.

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

Capítulo
13

# Análisis de supervivencia de Kaplan-Meier

Existen muchas situaciones en las se desea examinar la distribución de un período entre dos eventos, como la duración del empleo (tiempo transcurrido entre el contrato y el abandono de la empresa). Sin embargo, este tipo de datos incluye generalmente algunos casos censurados. Los casos censurados son casos para los que no se registra el segundo evento (por ejemplo, la gente que todavía está trabajando en la empresa al final del estudio). El procedimiento de Kaplan-Meier es un método de estimación de modelos hasta el evento en presencia de casos censurados. El modelo de Kaplan-Meier se basa en la estimación de las probabilidades condicionales en cada punto temporal cuando tiene lugar un evento y en tomar el límite del producto de esas probabilidades para estimar la tasa de supervivencia en cada punto temporal.

**Ejemplo.** ¿Posee algún beneficio terapéutico sobre la prolongación de la vida un nuevo tratamiento para el SIDA Se podría dirigir un estudio utilizando dos grupos de pacientes de SIDA, uno que reciba la terapia tradicional y otro que reciba el tratamiento experimental. Al construir un modelo de Kaplan-Meier a partir de los datos, se podrán comparar las tasas de supervivencia globales entre los dos grupos, para determinar si el tratamiento experimental representa una mejora con respecto a la terapia tradicional. Si desea obtener información más detallada, también es posible representar gráficamente las funciones de impacto o de supervivencia y compararlas visualmente.

**Estadísticos.** La tabla de supervivencia, que incluye el tiempo, el estado, la supervivencia acumulada y el error típico, los eventos acumulados y el número que permanece; la media y mediana del tiempo de supervivencia, con el error típico y el intervalo de confianza al 95%. Gráficos: supervivencia, impacto, log de la supervivencia y uno menos la supervivencia.

**Datos.** La variable de tiempo deberá ser continua, la variable de estado puede ser continua o categórica y las variables de estrato y de factor deberán ser categóricas.

**Supuestos.** Las probabilidades para el evento de interés deben depender solamente del tiempo transcurrido desde el evento inicial (se asume que son estables con respecto al tiempo absoluto). Es decir, los casos que se introducen en el estudio en horas diferentes (por ejemplo, pacientes que inician el tratamiento en horas diferentes) se deberían comportar de manera similar. Tampoco deben existir diferencias sistemáticas entre los casos censurados y los no censurados. Si, por ejemplo, muchos de los casos censurados son pacientes en condiciones más graves, los resultados pueden resultar sesgados.

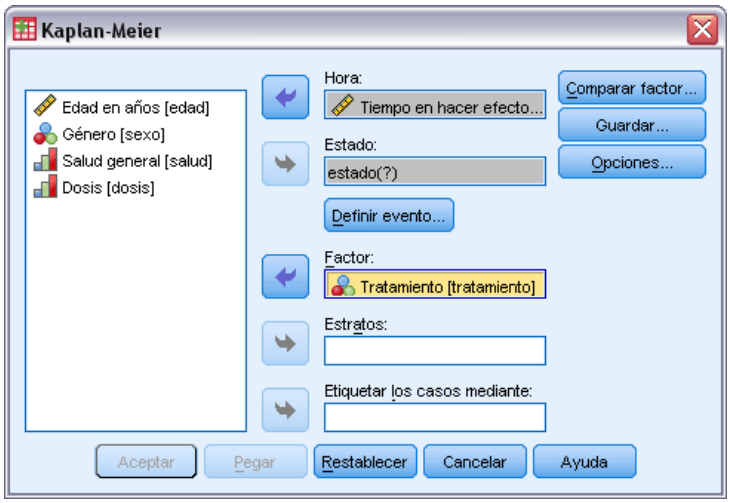
**Procedimientos relacionados.** El procedimiento de Kaplan-Meier utiliza un método de cálculo de las tablas de mortalidad que estima la función de impacto o supervivencia para el tiempo en que tiene lugar cada evento. El procedimiento Tablas de mortalidad utiliza un método actuarial al análisis de supervivencia que se basa en la partición del período de observación en intervalos de tiempo menores y puede ser útil para trabajar con grandes muestras. Si dispone de variables

que cree que están relacionadas con el tiempo de supervivencia o variables que desea controlar (covariables), utilice el procedimiento Regresión de Cox. Si las covariables pueden tener distintos valores en diferentes puntos temporales para el mismo caso, utilice el procedimiento Regresión de Cox con covariables dependientes del tiempo.

***P ara obtener un análisis de supervivencia de Kaplan-Meier***

- En los menús, seleccione:
  Analizar > Supervivencia > Kaplan-Meier...

Figura 13-1
*Cuadro de diálogo Kaplan-Meier*

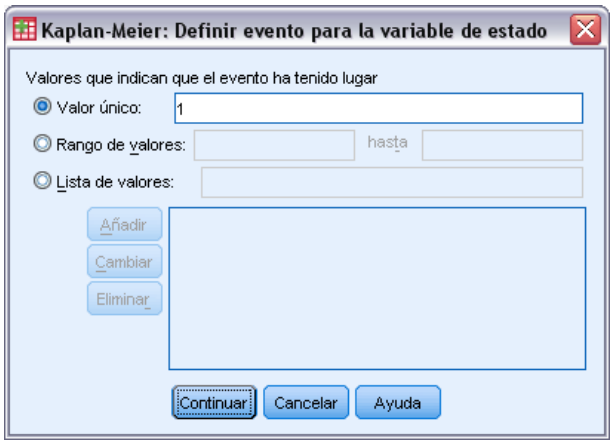



- Seleccione una variable de tiempo.
- Seleccione una variable de estado que identifique los casos para los que ha tenido lugar el evento terminal. Esta variable puede ser numérica o de **cadena corta**. A continuación, pulse en Definir evento.

  Si lo desea, puede seleccionar una variable de factor para examinar las diferencias entre grupos. Además es posible seleccionar una variable de estrato, que generará análisis diferentes para cada nivel (cada estrato) de la variable.

# Kaplan-Meier: Definir evento para la variable de estado

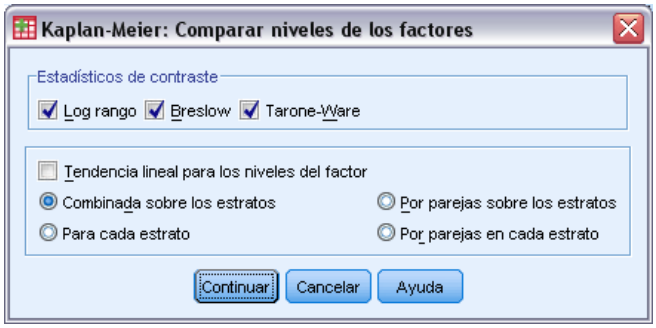
Figura 13-2
*Cuadro de diálogo Kaplan-Meier: Definir evento para la variable de estado*




Introduzca el valor o valores que indican que el evento terminal ha tenido lugar. Se puede introducir un solo valor, un rango de valores o una lista de valores. La opción Rango de valores solamente estará disponible si la variable de estado es numérica.

# Kaplan-Meier: Comparar niveles de los factores

Figura 13-3
*Cuadro de diálogo Kaplan-Meier: Comparar niveles de los factores*




Se pueden solicitar estadísticos para contrastar la igualdad de las distribuciones de supervivencia para los diferentes niveles del factor. Los estadísticos disponibles son Log rango, Breslow y Tarone-Ware. Seleccione una de las opciones para especificar las comparaciones que se van a realizar: Combinada sobre los estratos, Para cada estrato, Por parejas sobre los estratos o Por parejas en cada estrato.

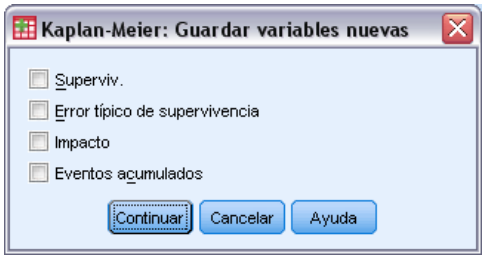
- **Log rango.** Prueba para contrastar la igualdad de las distribuciones de supervivencia. En esta prueba, todos los puntos del tiempo se ponderan por igual.
- **Breslow.** Prueba para contrastar la igualdad de las distribuciones de supervivencia. Los puntos del tiempo se ponderan por el número de los casos bajo riesgo que hay en cada punto del tiempo.

- **Tarone-Ware.** Prueba para contrastar la igualdad de las distribuciones de supervivencia. Los puntos del tiempo se multiplican por la raíz cuadrada del número de los casos bajo riesgo que hay en cada punto del tiempo.
- **Combinada sobre los estratos.** Compara todos los niveles del factor en una única prueba, para contrastar la igualdad de las curvas de supervivencia.
- **Por parejas sobre los estratos.** Compara cada par diferente de niveles del factor. No están disponibles las pruebas de tendencia por parejas.
- **Para cada estrato.** Realiza una prueba de igualdad para todos los niveles del factor, distinta para cada estrato. Si no tiene una variable de estratificación, las pruebas no se realizarán.
- **Por parejas en cada estrato.** Compara cada par diferente de niveles del factor en cada estrato. No están disponibles las pruebas de tendencia por parejas. Si no tiene una variable de estratificación, las pruebas no se realizarán.

**Tendencia lineal para los niveles del factor.** Permite contrastar la tendencia lineal a lo largo de los niveles del factor. Esta opción solamente estará disponible para las comparaciones globales (en vez de por parejas) de los niveles del factor.

## *Kaplan-Meier: Guardar variables nuevas*

Figura 13-4
*Cuadro de diálogo Kaplan-Meier: Guardar variables nuevas*



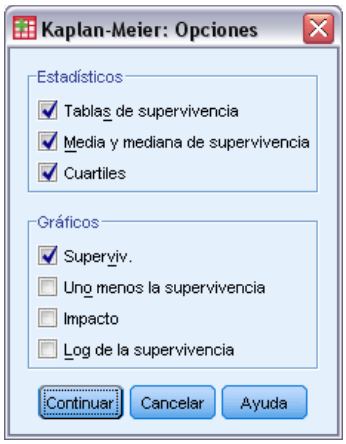

Es posible guardar información de la tabla de Kaplan-Meier como nuevas variables, información que se podrá utilizar en análisis subsiguientes para contrastar hipótesis o verificar los supuestos. Se pueden guardar estimaciones de la supervivencia, el error típico de la supervivencia, el impacto y los eventos acumulados, como nuevas variables.

- **Supervivencia.** Estimación de la probabilidad de supervivencia acumulada. El nombre de variable por defecto es el prefijo sur_ con un número secuencial. Por ejemplo, si sur_1 ya existe, Kaplan-Meier asigna el nombre de variable sur_2.
- **Error típico de supervivencia.** Error típico de la estimación de la supervivencia acumulada. El nombre de variable por defecto es el prefijo se_ con un número secuencial. Por ejemplo, si se_1 ya existe, Kaplan-Meier asigna el nombre de variable se_2.

- **Impacto.** Estimación de la función de impacto acumulada. El nombre de variable por defecto es el prefijo haz_ con un número secuencial. Por ejemplo, si haz_1 ya existe, Kaplan-Meier asigna el nombre de variable haz_2.
- **Eventos acumulados.** Frecuencia acumulada de los eventos, cuando los casos se ordenan por los tiempos de supervivencia y por los códigos de estado. El nombre de variable por defecto es el prefijo cum_ con un número secuencial. Por ejemplo, si cum_1 ya existe, Kaplan-Meier asigna el nombre de variable cum_2.

## Kaplan-Meier: Opciones

Figura 13-5
*Cuadro de diálogo Kaplan-Meier: Opciones*




Es posible solicitar varios tipos de resultados del análisis Kaplan-Meier.

**Estadísticos.** Es posible seleccionar que se muestren estadísticos para las funciones de supervivencia calculadas, incluyendo las tablas de supervivencia, la media y mediana de supervivencia y los cuartiles. Si se han incluido variables de factor, se generan estadísticos separados para cada grupo.

**Diagramas.** Permite examinar visualmente las funciones de supervivencia, uno menos la supervivencia, impacto y log de la supervivencia. Si se han incluido variables de factor, se representarán las funciones para cada grupo.

- **Supervivencia.** Muestra la función de supervivencia acumulada, en una escala lineal.
- **Uno menos la supervivencia.** Representa la función uno menos la supervivencia en una escala lineal.
- **Impacto.** Muestra la función de impacto acumulada, en una escala lineal.
- **Log de la supervivencia.** Muestra la función de supervivencia, en una escala logarítmica.

## Funciones adicionales del comando KM

Con el lenguaje de sintaxis de comandos también podrá:

- Obtener tablas de frecuencias que consideren los casos perdidos durante el seguimiento como una categoría diferente de los casos censurados.

- Especificar el espaciado desigual para la prueba de la tendencia lineal.
- Obtener percentiles diferentes a los cuartiles para la variable del tiempo de supervivencia.

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

Capítulo
14

# Análisis de regresión de Cox

La regresión de Cox genera un modelo predictivo para datos de tiempo de espera hasta el evento. El modelo genera una función de supervivencia que pronostica la probabilidad de que se haya producido el evento de interés en un momento dado *t* para determinados valores de las variables predictoras. La forma de la función de supervivencia y los coeficientes de regresión de los predictores se estiman mediante los sujetos observados; a continuación, se puede aplicar el modelo a los nuevos casos que tengan medidas para las variables predictoras. Observe que la información de los casos censurados, es decir, aquellos que no han experimentado el evento de interés durante el tiempo de observación, contribuye de manera útil a la estimación del modelo.

**Ejemplo.** ¿Corren los hombres y las mujeres diferentes riesgos de desarrollar cáncer de pulmón a causa del consumo de cigarrillos ? Construyendo un modelo de regresión de Cox, cuyas covariables sean el consumo diario de cigarrillos y el sexo, es posible contrastar las hipótesis sobre los efectos del consumo de tabaco y del sexo sobre el tiempo hasta el momento de la aparición de un cáncer de pulmón.

**Estadísticos.** Para cada modelo: *–2LL*, el estadístico de la razón de verosimilitud y el chi-cuadrado global. Para las variables dentro del modelo: Estimaciones de los parámetros, Errores típicos y Estadísticos de Wald. Para variables que no estén en el modelo: Estadísticos de puntuación y Chi-cuadrado residual.

**Datos.** La variable de tiempo debería ser cuantitativa, pero la variable de estado puede ser categórica o continua. Las variables independientes (las covariables) pueden ser continuas o categóricas; si son categóricas, deberán ser auxiliares (dummy) o estar codificadas con indicadores (existe una opción dentro del procedimiento para recodificar las variables categóricas automáticamente). Las variables de estratos deberían ser categóricas, codificadas como valores enteros o cadenas cortas.

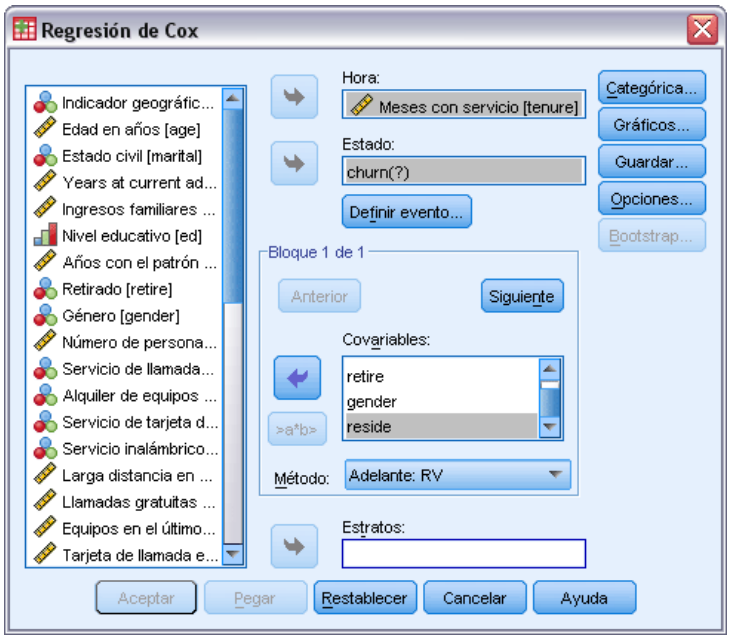
**Supuestos.** Las observaciones deben ser independientes y la tasa de impacto debe ser constante a lo largo del tiempo; es decir, la proporcionalidad de los impactos de un caso a otro no debe variar en función del tiempo. El último supuesto se conoce como el **supuesto de impactos proporcionales**.

**Procedimientos relacionados.** Si el supuesto de impactos proporcionales no se conserva (véase más arriba), es posible que deba utilizar el procedimiento de Cox con covariables dependientes del tiempo. Si no posee covariables o si solamente posee una covariable categórica, es posible utilizar las Tablas de mortalidad o el procedimiento de Kaplan-Meier para examinar las funciones de impacto o de supervivencia para las muestras. Si no posee datos censurados en la muestra (es decir, si todos los casos experimentaron el evento terminal), es posible utilizar el procedimiento Regresión lineal para modelar la relación entre las variables predictoras y el tiempo de espera hasta el evento.

***Para obtener un análisis de regresión de Cox***

▶ Seleccione en los menús:
Analizar > Superviv. > Regresión de Cox...

Figura 14-1
*Cuadro de diálogo Regresión de Cox*




▶ Seleccione una variable de tiempo. No se analizan aquellos casos en los que los valores del tiempo son negativos.

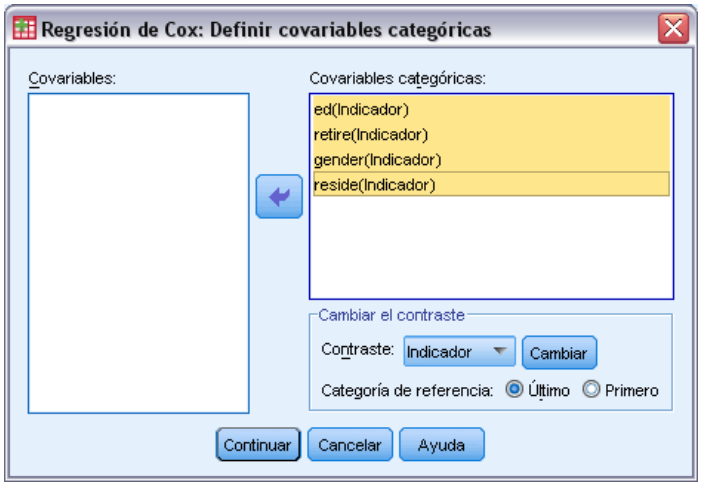
▶ Seleccione una variable de estado y pulse en Definir evento.

▶ Seleccione una o varias covariables. Para incluir términos de interacción, seleccione todas las variables contenidas en la interacción y pulse en >a*b>.

Si lo desea, es posible calcular modelos diferentes para diferentes grupos definiendo una variable para los estratos.

# Regresión de Cox: Definir variables categóricas

Figura 14-2
*Cuadro de diálogo Regresión de Cox: Definir variables categóricas*




Es posible especificar los detalles sobre cómo gestionará el procedimiento de Regresión de Cox las variables categóricas:

**Covariables.** Muestra 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.

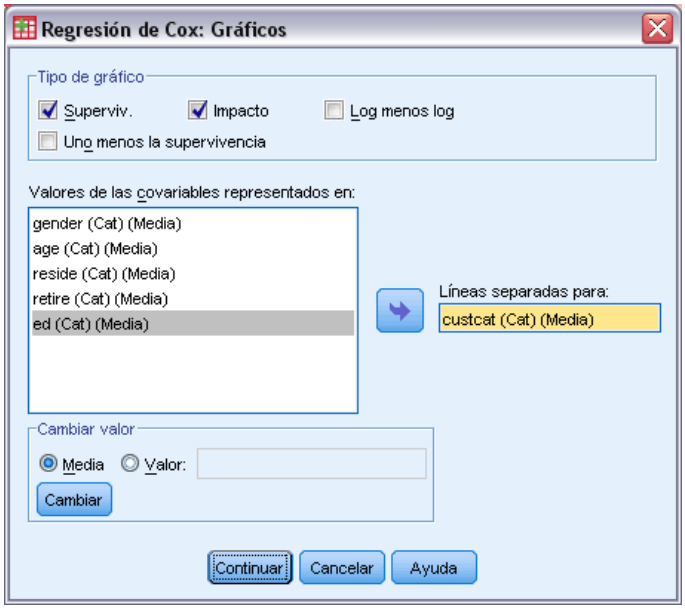
- **Polinómico.** Contrastes polinómicos ortogonales. Se supone que las categorías están espaciadas equidistantemente. Los contrastes polinómicos sólo están disponibles para variables numéricas.
- **Desviación.** Cada categoría del predictor, excepto la categoría de referencia, se compara con el efecto 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 de Cox: Gráficos

Figura 14-3
*Cuadro de diálogo Regresión de Cox: Gráficos*




Los gráficos pueden ayudarle a evaluar el modelo estimado e interpretar los resultados. Es posible representar gráficamente las funciones de supervivencia, de impacto, log-menos-log y uno menos la supervivencia.

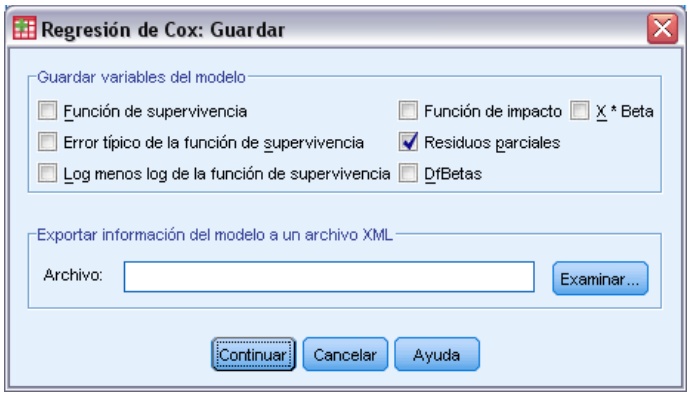
- **Superviv..** Muestra la función de supervivencia acumulada, en una escala lineal.
- **Impacto.** Muestra la función de impacto acumulada, en una escala lineal.
- **Log menos log.** La estimación de la supervivencia acumulada tras la transformación In(-ln) se aplica a la estimación.
- **Uno menos la supervivencia.** Representa la función uno menos la supervivencia en una escala lineal.

Como estas funciones dependen de los valores de las covariables, se deben utilizar valores constantes para las covariables con el fin de representar gráficamente las funciones respecto al tiempo. El valor por defecto es utilizar la media de cada covariable como un valor constante pero es posible introducir los propios valores para el gráfico utilizando el grupo de control Cambiar el valor.

Es posible representar gráficamente una línea diferente para cada valor de una covariable categórica, desplazando esa covariable al cuadro de texto Líneas separadas para. Esta opción solamente estará disponible para covariables categóricas, con la expresión (Cat) marcada después de sus nombres en la lista Valores de las covariables representados en.

## Regresión de Cox: Guardar nuevas variables

Figura 14-4
*Cuadro de diálogo Regresión de Cox: Guardar nuevas variables*




Es posible guardar varios resultados del análisis como nuevas variables. Estas variables se pueden utilizar en análisis siguientes para contrastar hipótesis o para comprobar supuestos.

**Guardar variables del modelo.** Permite guardar la función de supervivencia y su error típico, estimaciones de log menos log, función de impacto, residuos parciales, DfBeta(s) de la regresión y predictor lineal X*Beta como nuevas variables.

- **Función de supervivencia.** El valor de la función de supervivencia acumulada para un tiempo dado. Es igual a la probabilidad de supervivencia hasta ese período de tiempo.
- **Log menos log de la función de supervivencia.** La estimación de la supervivencia acumulada tras la transformación In(-ln) se aplica a la estimación.
- **Función de impacto.** Guarda la estimación de función de impacto acumulada (también llamado el residuo de Cox-Snell)
- **Residuos parciales.** Puede representar los residuos parciales respecto al tiempo para contrastar el supuesto de proporcionalidad de los impactos. Se guarda una variable por cada covariable en el modelo final. Los residuos parciales están disponibles sólo para los modelos que contienen al menos una covariable.

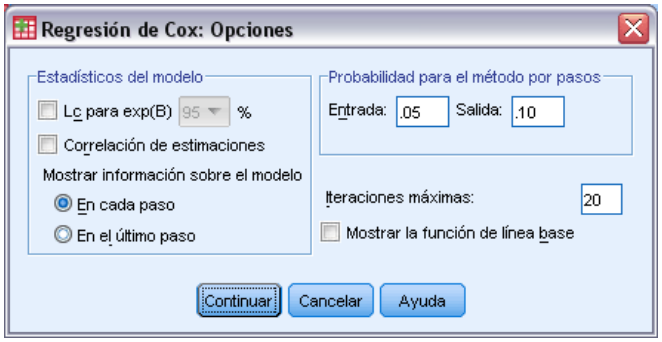
- **DfBetas.** Cambio estimado en un coeficiente si se elimina un caso. Se guarda una variable por cada covariable en el modelo final. Las DfBetas están disponibles sólo para los modelos que contienen al menos una covariable.
- **X*Beta.** Puntuación de la predicción lineal. La suma del producto de los valores de las covariables, centradas en la media (las puntuaciones diferenciales), por sus correspondientes parámetros estimados para cada caso.

Si está ejecutando Cox con una covariable dependiente del tiempo, DfBeta(s) y la variable predictora lineal X*Beta son las únicas variables que se pueden guardar.

**Exportar información del modelo a un archivo XML.** Las estimaciones de los parámetros se exportan al archivo especificado en formato XML. Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

# *Regresión de Cox: Opciones*

Figura 14-5
*Cuadro de diálogo Regresión de Cox: Opciones*




Es posible controlar diferentes aspectos del análisis y los resultados.

**Estadísticos del modelo.** Es posible obtener estadísticos para los parámetros del modelo, incluyendo los intervalos de confianza para exp(*B*) y la correlación de las estimaciones. Es posible solicitar estos estadísticos en cada paso o solamente en el último paso.

**Probabilidad para el método por pasos.** Si se ha seleccionado un método por pasos sucesivos, es posible especificar la probabilidad para la entrada o la exclusión desde el modelo. Una variable será introducida si el nivel de significación de su *F* para entrar es menor que el valor de Entrada y una variable será eliminada si el nivel de significación es mayor que el valor de Salida. El valor de Entrada debe ser menor que el valor de Salida.

**Nº máximo de iteraciones.** Permite especificar el número máximo de iteraciones para el modelo, que controla durante cuánto tiempo el procedimiento buscará una solución.

**Mostrar la función de línea base.** Permite visualizar la función de impacto basal y la supervivencia acumulada en la media de las covariables. Esta presentación no está disponible si se han especificado covariables dependientes del tiempo.

## Regresión de Cox: Definir evento para la variable de estado

Introduzca el valor o valores que indican que el evento terminal ha tenido lugar. Se puede introducir un solo valor, un rango de valores o una lista de valores. La opción Rango de valores solamente estará disponible si la variable de estado es numérica.

## Funciones adicionales del comando COXREG

La sintaxis de comandos también le permite:

- Obtener tablas de frecuencias que consideren los casos perdidos durante el seguimiento como una categoría diferente de los casos censurados.
- Seleccionar una categoría de referencia, que no sea la primera ni la última, para los métodos de contraste de indicador, simple y de desviación.
- Especificar un espaciado desigual entre las categorías para el método de contraste polinómico.
- Especificar los criterios de iteración adicionales.
- Controlar el tratamiento de los valores perdidos.
- Especificar los nombres para las variables guardadas.
- Guardar los resultados en un archivo de sistema IBM® SPSS® Statistics externo.
- Mantener los datos de cada grupo de segmentación del archivo en un archivo temporal externo durante el proceso. Esto puede contribuir a conservar los recursos de memoria cuando se ejecutan los análisis con grandes conjuntos de datos. No se encuentra disponible con covariables dependientes del tiempo.

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

Capítulo
# 15
# Calcular covariable dependiente del tiempo

Existen ciertas situaciones en las que interesa calcular un modelo de regresión de Cox, pero no se cumple el supuesto de tasas de impacto proporcionales. Es decir, que las tasas de impacto cambian con el tiempo; los valores de una (o de varias) de las covariables son diferentes en los distintos puntos temporales. En esos casos, es necesario utilizar un modelo de regresión de Cox extendido, que permita especificar **las covariables dependientes del tiempo**.

Con el fin de analizar dicho modelo, debe definir primero una covariable dependiente del tiempo (también se pueden especificar múltiples covariables dependientes del tiempo usando la sintaxis de comandos). (Estas covariables se pueden especificar usando la sintaxis de comandos). Para facilitar esta tarea cuenta con una variable del sistema que representa el tiempo. Esta variable se llama *T_*. Puede utilizar esta variable para definir covariables dependientes del tiempo empleando dos métodos generales:

- Si desea contrastar el supuesto de tasas de impacto proporcionales con respecto a una covariable particular o bien desea estimar un modelo de regresión de Cox extendido que permita impactos no proporcionales, defina la covariable dependiente del tiempo como una función de la variable de tiempo *T_* y la covariable en cuestión. Un ejemplo común sería el simple producto de la variable de tiempo y la covariable, pero también se pueden especificar funciones más complejas. La comparación de la significación del coeficiente de la covariable dependiente del tiempo le indicará si es razonable el supuesto de proporcionalidad de los impactos.
- Algunas variables pueden tener valores distintos en períodos diferentes del tiempo, pero no están sistemáticamente relacionadas con el tiempo. En tales casos es necesario definir una **covariable dependiente del tiempo segmentada**, lo cual puede llevarse a cabo usando las **expresiones lógicas**. Las expresiones lógicas toman el valor 1 cuando son verdaderas y el valor 0 cuando son falsas. Es posible crear una covariable dependiente del tiempo a partir de un conjunto de medidas, usando una serie de expresiones lógicas. Por ejemplo, si se toma la tensión una vez a la semana durante cuatro semanas (identificadas como *BP1* a *BP4*), puede definir el covariable dependiente del tiempo como (*T_* < 1) * *BP1* + (*T_* >= 1 & *T_* < 2) * *BP2* + (*T_* >= 2 & *T_* < 3) * *BP3* + (*T_* >= 3 & *T_* < 4) * *BP4*. Tenga en cuenta que exactamente uno de los términos entre paréntesis será igual a uno para cualquier caso dado y el resto serán todos 0. En otras palabras, esta función se puede interpretar diciendo que "Si el tiempo es inferior a una semana, use *BP1*; si es más de una semana pero menos de dos, utilice *BP2*, y así sucesivamente".

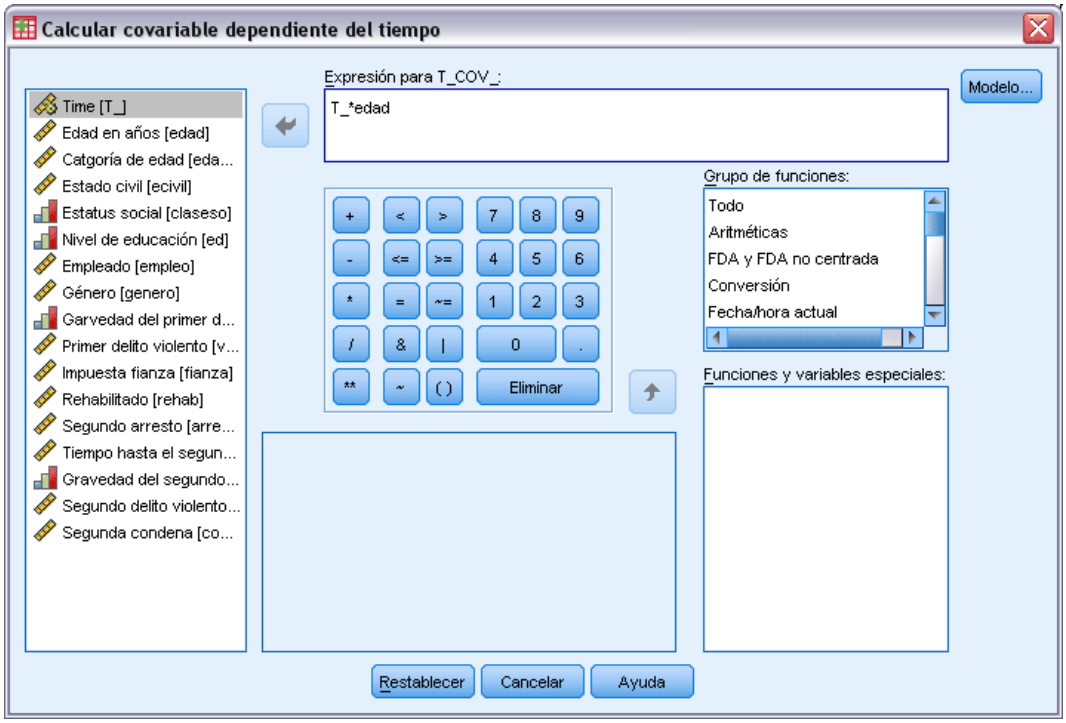
Puede utilizar los controles de generación de funciones para crear la expresión para la covariable dependiente del tiempo, o bien introducirla directamente en el área de texto Expresión para T_COV_. Tenga en cuenta que las constantes de cadena deben ir entre comillas o apóstrofes y que las constantes numéricas se deben escribir en formato americano, con el punto como

separador de la parte decimal. La variable resultante se llama *T_COV_* y se debe incluir como una covariable en el modelo de regresión de Cox.

## Para calcular una covariable dependiente del tiempo

- En los menús, seleccione:
  Analizar > Supervivencia > Cox con covariable dep. del tiempo...

Figura 15-1
*Cuadro de diálogo Calcular covariable dependiente del tiempo*




- Introduzca una expresión para la covariable dependiente del tiempo.
- Seleccione Modelo para continuar con la regresión de Cox.

*Nota*: Asegúrese de incluir la nueva variable *T_COV_* como covariable en el modelo de regresión de Cox.

Si desea obtener más información, consulte el tema Análisis de regresión de Cox en el capítulo 14 el p. 150.

## Regresión de Cox con covariables dependientes del tiempo: Funciones adicionales

El lenguaje de sintaxis de comandos permite especificar múltiples covariables dependientes del tiempo. Dispone además de otras funciones de sintaxis de comandos para la regresión de Cox con o sin covariables dependientes del tiempo.

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

Apéndice

A

# 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

| | | |
|---|---|---|
| ( 1/3 | 1/3 | 1/3 ) |
| ( 2/3 | –1/3 | –1/3 ) |
| ( –1/3 | –1/3 | 2/3 ) |

## *Simple*

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

| | | | | | |
|---|---|---|---|---|---|
| media | ( 1/*k* | 1/*k* | ··· | 1/*k* | 1/*k* ) |
| gl(1) | ( 1 | 0 | ··· | 0 | –1 ) |
| gl(2) | ( 0 | 1 | ··· | 0 | –1 ) |
| . | | . | | | |
| . | | . | | | |
| gl(k–1) | ( 0 | 0 | ··· | 1 | –1 ) |

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:

| | | | |
|---|---|---|---|
| ( 1/4 | 1/4 | 1/4 | 1/4 ) |
| ( 1 | 0 | 0 | –1 ) |
| ( 0 | 1 | 0 | –1 ) |
| ( 0 | 0 | 1 | –1 ) |

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

| | | | |
|---|---|---|---|
| ( 1/4 | 1/4 | 1/4 | 1/4 ) |
| ( 1 | –1 | 0 | 0 ) |
| ( 0 | –1 | 1 | 0 ) |
| ( 0 | –1 | 0 | 1 ) |

# Helmert

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

| | | | | | |
|---|---|---|---|---|---|
| media | ( 1/*k* | 1/*k* | ... | 1/*k* | 1/*k* ) |
| gl(1) | ( 1 | –1/(*k*–1) | ... | –1/(*k*–1) | –1/(*k*–1) ) |
| gl(2) | ( 0 | 1 | ... | –1/(*k*–2) | –1/(*k*–2) ) |
| . | | . | | | |
| . | | . | | | |
| gl(k–2) | ( 0 | 0 | 1 | –1/2 | –1/2 |
| gl(k–1) | ( 0 | 0 | ... | 1 | –1 ) |

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:

| | | | |
|---|---|---|---|
| ( 1/4 | 1/4 | 1/4 | 1/4 ) |
| ( 1 | –1/3 | –1/3 | –1/3 ) |
| ( 0 | 1 | –1/2 | –1/2 ) |
| ( 0 | 0 | 1 | –1 ) |

# 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

| | | | | | |
|---|---|---|---|---|---|
| media | ( 1/*k* | 1/*k* | 1/*k* | ... | 1/*k* ) |
| gl(1) | ( –1 | 1 | 0 | ... | 0 ) |
| gl(2) | ( –1/2 | –1/2 | 1 | ... | 0 ) |
| . | | . | | | |
| . | | . | | | |
| gl(k–1) | ( –1/(*k*–1) | –1/(*k*–1) | –1/(*k*–1) | ... | 1 ) |

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:

| | | | |
|---|---|---|---|
| ( 1/4 | 1/4 | 1/4 | 1/4 ) |
| ( –1 | 1 | 0 | 0 ) |
| ( –1/2 | –1/2 | 1 | 0 ) |
| ( –1/3 | –1/3 | –1/3 | 1 ) |

## 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:

| | | | |
|---|---|---|---|
| ( 1/4 | 1/4 | 1/4 | 1/4 ) |
| ( 1 | –1 | 0 | 0 ) |
| ( 0 | 1 | –1 | 0 ) |
| ( 0 | 0 | 1 | –1 ) |

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 | 1 | 1 | 1 ) | ponderaciones para el cálculo de la media |
| ( 3 | –1 | –1 | –1 ) | compare 1° con 2° hasta 4° |
| ( 0 | 2 | –1 | –1 ) | compare 2° con 3° y 4° |
| ( 0 | 0 | 1 | –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:

| | |
|---|---|
| Filas 2 y 3: | (3)(0) + (–1)(2) + (–1)(–1) + (–1)(–1) = 0 |
| Filas 2 y 4: | (3)(0) + (–1)(0) + (–1)(1) + (–1)(–1) = 0 |
| Filas 3 y 4: | (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.

Apéndice

B

# Estructuras de covarianza

Esta sección ofrece información adicional sobre las estructuras de covarianza.

**Dependencia Ante: Primer orden.** Esta estructura de covarianza tiene varianzas heterogéneas y correlaciones heterogéneas entre los elementos adyacentes. La correlación entre dos elementos que no son adyacentes es el producto de las correlaciones entre los elementos que se encuentran entre los elementos de interés.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3 \\ \sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}$$

**AR(1).** Se trata de una estructura autorregresiva de primer orden con varianzas homogéneas. La correlación entre dos elementos es igual a rho en el caso de elementos adyacentes, $\rho^2$ cuando se trata de elementos separados entre sí por un tercero, y así sucesivamente. $\rho$ está limitado de manera que $-1<\rho<1$.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

**AR(1):** Heterogénea. Se trata de una estructura autorregresiva de primer orden con varianzas heterogéneas. La correlación entre dos elementos es igual a $\rho$ en el caso de elementos adyacentes, $\rho^2$ cuando se trata de dos elementos separados entre sí por un tercero, y así sucesivamente. $\rho$ está limitado y debe estar comprendido entre –1 y 1.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

**ARMA(1,1).** Se trata de una estructura de media móvil autorregresiva. Tiene varianzas homogéneas. La correlación entre dos elementos es igual a $\phi*\rho$ en el caso de elementos adyacentes, $\phi*(\rho^2)$ en el caso de elementos separados por un tercero, y así sucesivamente. $\rho$ y $\phi$ son los parámetros autorregresivo y de media móvil, respectivamente, y sus valores deben estar comprendidos entre –1 y 1 (ambos incluidos).

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi\rho^2 \\ \phi\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & 1 \end{bmatrix}$$

**Simetría compuesta.** Esta estructura tiene una varianza y una covarianza constantes.

$$\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1^2 \end{bmatrix}$$

**Simetría compuesta: Métrica de correlación.** Esta estructura de covarianza tiene varianzas y correlaciones homogéneas entre los elementos.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

**Simetría compuesta:** Heterogénea. Esta estructura de covarianza tiene varianzas heterogéneas y una correlación constante entre los elementos.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho & \sigma_4\sigma_1\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

**Diagonal.** Esta estructura de covarianza tiene varianzas heterogéneas y una correlación cero entre los elementos.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

**Factor analítico: Primer orden.** Esta estructura de covarianza tiene varianzas heterogéneas que están compuestas de un término que es heterogéneo en los elementos y un término que es homogéneo en los elementos. La covarianza entre dos elementos es la raíz cuadrada del producto de sus términos de varianza heterogéneos.

$$\begin{bmatrix} \lambda_1^2+d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2+d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+d & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2+d \end{bmatrix}$$

**Factor analítico: Primer orden, Heterogéneo.** Esta estructura de covarianza tiene varianzas heterogéneas que están compuestas de dos términos que son heterogéneos en los elementos. La covarianza entre dos elementos es la raíz cuadrada del producto del primero de sus términos de varianza heterogéneos.

$$\begin{bmatrix} \lambda_1^2 + d_1 & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2 + d_2 & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d_3 & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d_4 \end{bmatrix}$$

**Huynh-Feldt.** Se trata de una matriz "circular" en la que la covarianza entre dos elementos es igual a la media de las varianzas menos una constante. Ni las varianzas ni las covarianzas son constantes.

$$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2+\sigma_2^2}{2} - \lambda & \frac{\sigma_1^2+\sigma_3^2}{2} - \lambda & \frac{\sigma_1^2+\sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2+\sigma_2^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2+\sigma_3^2}{2} - \lambda & \frac{\sigma_2^2+\sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2+\sigma_3^2}{2} - \lambda & \frac{\sigma_2^2+\sigma_3^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2+\sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2+\sigma_4^2}{2} - \lambda & \frac{\sigma_2^2+\sigma_4^2}{2} - \lambda & \frac{\sigma_3^2+\sigma_4^2}{2} - \lambda & \sigma_4^2 \end{bmatrix}$$

**Identidad escalada.** Esta estructura tiene una varianza constante. Se asume que no existe correlación alguna entre los elementos.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

**Toeplitz.** Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre los elementos. La correlación entre elementos adyacentes es homogénea en pares de elementos adyacentes. La correlación entre elementos separados por un tercero vuelve a ser homogénea, y así sucesivamente.

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

**Toeplitz:** Heterogénea. Esta estructura de covarianza tiene varianzas heterogéneas y correlaciones heterogéneas entre los elementos. La correlación entre elementos adyacentes es homogénea en pares de elementos adyacentes. La correlación entre elementos separados por un tercero vuelve a ser homogénea, y así sucesivamente.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_1 & \sigma_4\sigma_2\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_4\sigma_3\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}$$

**Sin estructura.** Es una matriz de covarianzas completamente general.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

**Sin estructura: Métrica de correlación.** Esta estructura de covarianza tiene varianzas heterogéneas y correlaciones heterogéneas.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$$

**Componentes de la varianza.** Esta estructura asigna una estructura de identidad escalada (ID) a cada uno de los efectos aleatorios especificados.

Apéndice
C

# 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 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.shmtl*.

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.

# Índice


análisis de covarianza
  en MLG multivariante, 2
análisis de la varianza
  en los componentes de la varianza, 34
análisis de supervivencia
  en Kaplan-Meier, 144
  en la regresión de Cox, 150
  en las tablas de mortalidad, 139
  Regresión de Cox dependiente del tiempo, 157
análisis loglineal, 123
  Análisis loglineal general, 127
  Análisis loglineal logit, 133
Análisis loglineal general
  almacenamiento de valores pronosticados, 131
  almacenamiento de variables, 131
  contrastes, 127
  covariables de casilla, 127
  criterios, 130
  distribución de recuentos de casillas, 127
  especificación de modelo, 129
  estructuras de casilla, 127
  factores, 127
  funciones adicionales del comando, 131
  gráficos, 130
  intervalos de confianza, 130
  opciones de presentación, 130
  residuos, 131
Análisis loglineal logit, 133
  almacenamiento de variables, 137
  contrastes, 133
  covariables de casilla, 133
  criterios, 136
  distribución de recuentos de casillas, 133
  especificación de modelo, 135
  estructuras de casilla, 133
  factores, 133
  gráficos, 136
  intervalos de confianza, 136
  opciones de presentación, 136
  residuos, 137
  valores pronosticados, 137
Análisis loglineal: Selección de modelo, 123
  definición de los rangos del factor, 124
  funciones adicionales del comando, 126
  modelos, 125
  opciones, 126
ANOVA
  en MLG medidas repetidas, 15
  en MLG multivariante, 2
ANOVA multivariada, 2

bondad de ajuste
  en ecuaciones de estimación generalizadas, 92
  en modelos lineales generalizados, 65
Bonferroni
  en MLG medidas repetidas, 24
  en MLG multivariante, 9

*C* de Dunnett
  en MLG medidas repetidas, 24
  en MLG multivariante, 9
casos censurados
  en Kaplan-Meier, 144
  en la regresión de Cox, 150
  en las tablas de mortalidad, 139
categoría de referencia
  en ecuaciones de estimación generalizadas, 81, 84
  en modelos lineales generalizados, 56
clase generadora
  en el análisis loglineal de selección de modelo, 125
construcción de términos, 5, 21, 33, 125, 129, 136
contraste de multiplicador de Lagrange
  en modelos lineales generalizados, 65
contrastes
  en el análisis loglineal general, 127
  en el análisis loglineal logit, 133
  en la regresión de Cox, 152
convergencia de los parámetros
  en ecuaciones de estimación generalizadas, 87
  en modelos lineales generalizados, 61
  en modelos lineales mixtos, 45
convergencia del logaritmo de la verosimilitud
  en ecuaciones de estimación generalizadas, 87
  en modelos lineales generalizados, 61
  en modelos lineales mixtos, 45
Convergencia hessiana
  en ecuaciones de estimación generalizadas, 87
  en modelos lineales generalizados, 61
covariables
  en la regresión de Cox, 152
covariables de cadena
  en la regresión de Cox, 152
covariables segmentadas dependientes del tiempo
  en la regresión de Cox, 157

descomposición jerárquica, 5, 21
  en los componentes de la varianza, 35
desviación típica
  en MLG medidas repetidas, 28
  en MLG multivariante, 12

diagramas de dispersión por nivel
en MLG medidas repetidas, 28
en MLG multivariante, 12
diferencia honestamente significativa de Tukey
en MLG medidas repetidas, 24
en MLG multivariante, 9
diferencia menos significativa
en MLG medidas repetidas, 24
en MLG multivariante, 9
Distancia de Cook
en MLG, 11
en MLG medidas repetidas, 26
en modelos lineales generalizados, 69
distribución binomial
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución binomial negativa
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución De Gauss inversa
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución de Poisson
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución gamma
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución multinomial
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución normal
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
distribución Tweedie
en ecuaciones de estimación generalizadas, 78
en modelos lineales generalizados, 52
DMS de Fisher
en MLG medidas repetidas, 24
en MLG multivariante, 9

Ecuaciones de estimación generalizadas, 73
categoría de referencia para respuesta binaria, 81
criterios de estimación, 87
especificación de modelo, 85
estadísticos, 91
exportación del modelo, 97
guardar variables en el conjunto de datos activo, 95
medias marginales estimadas, 93
opciones para factores categóricos, 84
predictores, 83
respuesta, 80
tipo de modelo, 77
valores iniciales, 89
efectos aleatorios
en modelos lineales mixtos, 43
efectos fijos
en modelos lineales mixtos, 41
eliminación hacia atrás
en el análisis loglineal de selección de modelo, 123
EMNCI
en los componentes de la varianza, 34
error típico
en MLG, 11
en MLG medidas repetidas, 26, 28
en MLG multivariante, 12
estadístico de Wald
en el análisis loglineal general, 127
en el análisis loglineal logit, 133
estadísticos descriptivos
en ecuaciones de estimación generalizadas, 92
en MLG medidas repetidas, 28
en MLG multivariante, 12
en modelos lineales generalizados, 65
en modelos lineales mixtos, 46
estimación de la máxima verosimilitud
en los componentes de la varianza, 34
estimación de la máxima verosimilitud restringida
en los componentes de la varianza, 34
estimaciones de los parámetros
en ecuaciones de estimación generalizadas, 92
en el análisis loglineal de selección de modelo, 126
en el análisis loglineal general, 127
en el análisis loglineal logit, 133
en MLG medidas repetidas, 28
en MLG multivariante, 12
en modelos lineales generalizados, 65
en modelos lineales mixtos, 46
estimaciones de potencia
en MLG medidas repetidas, 28
en MLG multivariante, 12
estimaciones de tamaño de efecto
en MLG medidas repetidas, 28
en MLG multivariante, 12
estructuras de covarianza, 166
en modelos lineales mixtos, 166
eta-cuadrado
en MLG medidas repetidas, 28
en MLG multivariante, 12

*F* múltiple de Ryan-Einot-Gabriel-Welsch
en MLG medidas repetidas, 24
en MLG multivariante, 9
factores
en MLG medidas repetidas, 18
frecuencias
en el análisis loglineal de selección de modelo, 126
función de enlace
modelos mixtos lineales generalizados, 103
función de enlace binomial negativa
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53

función de enlace Cauchit acumulada
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace complementaria log
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace de identidad
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace de potencia
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace de potencia de las ventajas
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace log
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace log-log complementaria
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace log-log complementaria acumulada
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace log-log negativa
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace log-log negativa acumulada
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace logit
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace logit acumulada
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace probit
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de enlace probit acumulada
en ecuaciones de estimación generalizadas, 79
en modelos lineales generalizados, 53
función de supervivencia
en las tablas de mortalidad, 139
función estimable general
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65

GLOR
en el análisis loglineal general, 127
gráficos
en el análisis loglineal general, 130
en el análisis loglineal logit, 136
gráficos de los residuos
en MLG medidas repetidas, 28
en MLG multivariante, 12
gráficos de perfil
en MLG medidas repetidas, 23
en MLG multivariante, 8
gráficos de probabilidad normal
en el análisis loglineal de selección de modelo, 126
GT2 de Hochberg
en MLG medidas repetidas, 24
en MLG multivariante, 9

historial de iteraciones
en modelos lineales mixtos, 45
histórico de iteraciones
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65

información de los niveles del factor
en modelos lineales mixtos, 46
información del modelo
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65
intervalos de confianza
en el análisis loglineal general, 130
en el análisis loglineal logit, 136
en MLG medidas repetidas, 28
en MLG multivariante, 12
en modelos lineales mixtos, 46
iteraciones
en ecuaciones de estimación generalizadas, 87
en el análisis loglineal de selección de modelo, 126
en modelos lineales generalizados, 61

Kaplan-Meier, 144
almacenamiento de nuevas variables, 147
comparación de niveles del factor, 146
cuartiles, 148
definición de eventos, 146
ejemplo, 144
estadísticos, 144, 148
funciones adicionales del comando, 148
gráficos, 148
media y mediana de tiempos de supervivencia, 148
tablas de supervivencia, 148
tendencia lineal para los niveles del factor, 146
variables de estado de supervivencia, 146

legal notices, 170
log-razón de las ventajas generalizadas
en el análisis loglineal general, 127

matriz de correlaciones
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65
en modelos lineales mixtos, 46

matriz de covarianzas
en ecuaciones de estimación generalizadas, 87, 92
en MLG, 11
en modelos lineales generalizados, 61, 65
en modelos lineales mixtos, 46
matriz de covarianzas de efectos aleatorios
en modelos lineales mixtos, 46
matriz de covarianzas de los parámetros
en modelos lineales mixtos, 46
matriz de covarianzas residuales
en modelos lineales mixtos, 46
matriz de los coeficientes de los contrastes
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65
matriz L
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65
medias marginales estimadas
en ecuaciones de estimación generalizadas, 93
en MLG medidas repetidas, 28
en MLG multivariante, 12
en modelos lineales generalizados, 66
en modelos lineales mixtos, 47
medias observadas
en MLG medidas repetidas, 28
en MLG multivariante, 12
método de Newton-Raphson
en el análisis loglineal general, 127
en el análisis loglineal logit, 133
MLG
almacenamiento de matrices, 11
almacenamiento de variables, 11
MLG Medidas repetidas, 15
almacenamiento de variables, 26
contrastes post hoc, 24
definir factores, 18
diagnósticos, 28
funciones adicionales del comando, 29
gráficos de perfil, 23
medias marginales estimadas, 28
modelo, 20
opciones, 28
presentación, 28
MLG multivariante, 2
MLG Multivariante, 2, 14
contrastes post hoc, 9
covariables, 2
diagnósticos, 12
factores, 2
gráficos de perfil, 8
medias marginales estimadas, 12
opciones, 12
presentación, 12
variable dependiente, 2
modelo de impactos proporcionales
en la regresión de Cox, 150
modelos factoriales completos
en los componentes de la varianza, 33
en MLG medidas repetidas, 20
Modelos lineales generalizados, 50
categoría de referencia para respuesta binaria, 56
criterios de estimación, 61
distribución, 50
especificación de modelo, 59
estadísticos, 64
exportación del modelo, 70
función de enlace, 50
guardar variables en el conjunto de datos activo, 68
medias marginales estimadas, 66
opciones para factores categóricos, 58
predictores, 57
respuesta, 55
tipos de modelos, 50
valores iniciales, 63
Modelos lineales mixtos, 37, 166
almacenamiento de variables, 48
construcción de términos, 41–42
criterios de estimación, 45
efectos aleatorios, 43
efectos fijos, 41
estructura de covarianza, 166
funciones adicionales del comando, 49
medias marginales estimadas, 47
modelo, 46
términos de interacción, 41
modelos logit multinomiales, 133
modelos loglineales jerárquicos, 123
modelos mixtos
lineales, 37
modelos mixtos lineales generalizados, 100
anotaciones, 118–120
bloque de efectos aleatorios, 110
desplazamiento, 112
distribución de destino, 103
efectos aleatorios, 109
efectos fijos, 106
estructura de datos, 118
exportación del modelo, 116
función de enlace, 103
guardar campos, 116
medias estimadas, 121
medias marginales estimadas, 114
ponderación de análisis, 112
predicho por observado, 118
resumen de modelo, 117
tabla de clasificación, 118
términos personalizados, 107
vista de modelo, 117
modelos personalizados
en el análisis loglineal de selección de modelo, 125
en los componentes de la varianza, 33
en MLG medidas repetidas, 20

modelos saturados
en el análisis loglineal de selección de modelo, 125

Newman-Keuls
en MLG medidas repetidas, 24
en MLG multivariante, 9

parámetro de escala
en ecuaciones de estimación generalizadas, 87
en modelos lineales generalizados, 61
previas de los efectos aleatorios
en los componentes de la varianza, 34
productos cruzados
matrices de hipótesis y error, 12
prueba *b* de Tukey
en MLG medidas repetidas, 24
en MLG multivariante, 9
prueba de Breslow
en Kaplan-Meier, 146
Prueba de comparación por parejas de Gabriel
en MLG medidas repetidas, 24
en MLG multivariante, 9
Prueba de comparación por parejas de Games y Howell
en MLG medidas repetidas, 24
en MLG multivariante, 9
prueba de esfericidad de Bartlett
en MLG multivariante, 12
prueba de esfericidad de Mauchly
en MLG medidas repetidas, 28
prueba de Gehan
en las tablas de mortalidad, 142
prueba de Levene
en MLG medidas repetidas, 28
en MLG multivariante, 12
prueba de log rango
en Kaplan-Meier, 146
prueba de parámetros de covarianza
en modelos lineales mixtos, 46
prueba de rangos múltiples de Duncan
en MLG medidas repetidas, 24
en MLG multivariante, 9
prueba de Scheffé
en MLG medidas repetidas, 24
en MLG multivariante, 9
prueba de Tarone-Ware
en Kaplan-Meier, 146
prueba de Wilcoxon
en las tablas de mortalidad, 142
Prueba M de Box
en MLG multivariante, 12
Prueba t
en MLG medidas repetidas, 28
en MLG multivariante, 12
prueba *t* de Dunnett
en MLG medidas repetidas, 24
en MLG multivariante, 9
prueba *t* de Sidak
en MLG medidas repetidas, 24
en MLG multivariante, 9
prueba *t* de Waller-Duncan
en MLG medidas repetidas, 24
en MLG multivariante, 9
pruebas de homogeneidad de las varianzas
en MLG medidas repetidas, 28
en MLG multivariante, 12
puntuación
en modelos lineales mixtos, 45
puntuación de Fisher
en modelos lineales mixtos, 45

R-E-G-W *F*
en MLG medidas repetidas, 24
en MLG multivariante, 9
R-E-G-W *Q*
en MLG medidas repetidas, 24
en MLG multivariante, 9
rango múltiple de Ryan-Einot-Gabriel-Welsch
en MLG medidas repetidas, 24
en MLG multivariante, 9
razón de ventajas
en el análisis loglineal general, 127
Regresión de Cox, 150
almacenamiento de nuevas variables, 154
contrastes, 152
covariables, 150
covariables categóricas, 152
covariables de cadena, 152
covariables dependientes del tiempo, 157–158
definición de eventos, 156
DfBetas, 154
ejemplo, 150
entrada o exclusión por pasos, 155
estadísticos, 150, 155
función de impacto, 154
función de supervivencia, 154
funciones adicionales del comando, 156
funciones de línea base, 155
gráficos, 153
iteraciones, 155
residuos parciales, 154
variable del estado de supervivencia, 156
Regresión de Poisson
en el análisis loglineal general, 127
regresión multivariada, 2
residuo SCPC
en MLG medidas repetidas, 28
en MLG multivariante, 12
residuos
en ecuaciones de estimación generalizadas, 96
en el análisis loglineal de selección de modelo, 126
en el análisis loglineal general, 131
en el análisis loglineal logit, 137
en modelos lineales generalizados, 69

en modelos lineales mixtos, 48
residuos de desvianza
en modelos lineales generalizados, 69
residuos de Pearson
en ecuaciones de estimación generalizadas, 96
en modelos lineales generalizados, 69
residuos de verosimilitud
en modelos lineales generalizados, 69
residuos eliminados
en MLG, 11
en MLG medidas repetidas, 26
residuos no tipificados
en MLG, 11
en MLG medidas repetidas, 26
residuos tipificados
en MLG, 11
en MLG medidas repetidas, 26
resumen de procesamiento de casos
en ecuaciones de estimación generalizadas, 92
en modelos lineales generalizados, 65

SCPC
en MLG medidas repetidas, 28
en MLG multivariante, 12
separación
en ecuaciones de estimación generalizadas, 87
en modelos lineales generalizados, 61
Student-Newman-Keuls
en MLG medidas repetidas, 24
en MLG multivariante, 9
subdivisión por pasos
en ecuaciones de estimación generalizadas, 87
en modelos lineales generalizados, 61
en modelos lineales mixtos, 45
suma de cuadrados, 5, 21
en los componentes de la varianza, 35
en modelos lineales mixtos, 42
matrices de hipótesis y error, 12

T2 de Tamhane
en MLG medidas repetidas, 24
en MLG multivariante, 9
T3 de Dunnett
en MLG medidas repetidas, 24
en MLG multivariante, 9
tabla de contingencia
en el análisis loglineal de selección de modelo, 123
tablas de contingencia
en el análisis loglineal general, 127
Tablas de mortalidad, 139
comparación de niveles del factor, 142
ejemplo, 139
estadísticos, 139
función de supervivencia, 139
funciones adicionales del comando, 142
gráficos, 142
prueba de Wilcoxon (Gehan), 142
supresión de la presentación de tablas, 142
tasa de impacto, 139
variables de estado de supervivencia, 141
variables del factor, 141
tasa de impacto
en las tablas de mortalidad, 139
términos anidados
en ecuaciones de estimación generalizadas, 85
en modelos lineales generalizados, 59
en modelos lineales mixtos, 42
términos de interacción, 5, 21, 33, 125, 129, 136
en modelos lineales mixtos, 41
tolerancia para la singularidad
en modelos lineales mixtos, 45
trademarks, 171

valores de influencia
en MLG, 11
en MLG medidas repetidas, 26
en modelos lineales generalizados, 69
valores pronosticados
en el análisis loglineal general, 131
en el análisis loglineal logit, 137
en modelos lineales mixtos, 48
valores pronosticados fijos
en modelos lineales mixtos, 48
valores pronosticados ponderados
en MLG, 11
en MLG medidas repetidas, 26
variables de medidas repetidas
en modelos lineales mixtos, 39
variables de sujetos
en modelos lineales mixtos, 39
Variance Components, 31
almacenamiento de resultados, 36
funciones adicionales del comando, 36
modelo, 33
opciones, 34
vista de modelo
en modelos mixtos lineales generalizados, 117