

IBM SPSS Advanced Statistics 29

IBM

Nota

Antes de utilizar esta información y el producto al que se refiere, lea la información en [“Avisos” en la página 125](#).

Información del producto

Esta edición se aplica a la versión 29, release 0, modificación 0 de IBM® SPSS Statistics y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

© Copyright International Business Machines Corporation .

Contenido

Capítulo 1. Estadísticos avanzados.....	1
Introducción a Estadísticas avanzadas.....	1
Análisis GLM multivariante.....	1
Modelo MLG multivariante.....	3
MLG Multivariante: Contrastes.....	4
Gráficos de perfil de MLG multivariante.....	5
MLG multivariante: Comparaciones post hoc.....	5
Medias marginales estimadas GLM.....	6
Guardar MLG.....	7
MLG Multivariante: Opciones.....	8
Características adicionales del mandato GLM.....	9
MLG Medidas repetidas.....	9
MLG Medidas repetidas: Definir factores.....	12
MLG Medidas repetidas: Modelo.....	12
MLG Medidas repetidas: Contrastes.....	13
MLG Medidas repetidas: Gráficos de perfil.....	14
MLG Medidas repetidas: Comparaciones post hoc.....	15
Medias marginales estimadas GLM.....	16
MLG Medidas repetidas: guardar.....	16
GLM Medidas repetidas: Opciones.....	17
Características adicionales del mandato GLM.....	18
Análisis de componentes de la varianza.....	19
Componentes de la varianza: Modelo.....	20
Componentes de la varianza: Opciones.....	20
Componentes de la varianza: Guardar en archivo nuevo.....	21
Características adicionales del comando VARCOMP.....	22
Modelos lineales mixtos.....	22
Modelos lineales mixtos: Sujetos y Repetidos.....	23
Efectos fijos de los Modelos lineales mixtos.....	25
Efectos aleatorios de los Modelos lineales mixtos.....	26
Estimación de los Modelos lineales mixtos.....	27
Estadísticos de Modelos lineales mixtos.....	28
Medias marginales estimadas de modelos lineales mixtos.....	28
Guardar Modelos lineales mixtos.....	29
Características adicionales del comando MIXED.....	29
Modelos lineales generalizados.....	29
Modelos lineales generalizados: Respuesta.....	32
Modelos lineales generalizados: Predictores.....	33
Modelos lineales generalizados: Modelo.....	34
Modelos lineales generalizados: Estimación.....	34
Modelos lineales generalizados: Estadísticos.....	36
Modelos lineales generalizados: Medias marginales estimadas.....	37
Modelos lineales generalizados: Guardar.....	38
Modelos lineales generalizados: Exportar.....	39
Características adicionales del comando GENLIN.....	40
Ecuaciones de estimación generalizadas.....	40
Ecuaciones de estimación generalizadas: Tipo de modelo.....	42
Respuesta de las ecuaciones de estimación generalizadas.....	44
Ecuaciones de estimación generalizadas: Predictores.....	45
Ecuaciones de estimación generalizadas: Modelo.....	46
Ecuaciones de estimación generalizadas: Estimación.....	46

Ecuaciones de estimación generalizadas: Estadísticos.....	48
Ecuaciones de estimación generalizadas: Medias marginales estimadas.....	49
Ecuaciones de estimación generalizadas: Guardar.....	50
Ecuaciones de estimación generalizadas: Exportar.....	51
Características adicionales del comando GENLIN.....	52
Modelos mixtos lineales generalizados.....	52
Obtención de un modelo mixto lineal generalizado.....	54
Objetivo	54
Efectos fijos	56
Efectos aleatorios	58
Ponderación y desplazamiento	59
Opciones de generación generales	59
Estimación	60
Medias estimadas	60
Guardar	61
Vista de modelo	62
Análisis loglineal: Selección de modelo.....	66
Análisis loglineal: Definir rango.....	67
Análisis loglineal: Modelo.....	67
Análisis loglineal: Opciones.....	68
Características adicionales del comando HILOGLINEAR.....	68
Análisis loglineal general.....	68
Análisis loglineal general: Modelo.....	69
Análisis loglineal general: Opciones.....	70
Análisis loglineal general: Guardar.....	70
Características adicionales del comando GENLOG.....	71
Análisis loglineal logit.....	71
Análisis loglineal logit: Modelo.....	72
Análisis loglineal logit: Opciones.....	73
Análisis loglineal logit: Guardar.....	73
Características adicionales del comando GENLOG.....	74
Tablas de mortalidad.....	74
Tablas de mortalidad: Definir evento para la variable de estado.....	75
Tablas de mortalidad: Definir rango.....	76
Tablas de mortalidad: Opciones.....	76
Características adicionales del comando SURVIVAL.....	76
Análisis de supervivencia de Kaplan-Meier.....	76
Kaplan-Meier: Definir evento para la variable de estado.....	77
Kaplan-Meier: Comparar niveles de los factores.....	77
Kaplan-Meier: Guardar variables nuevas.....	78
Kaplan-Meier: Opciones.....	78
Características adicionales del comando KM.....	78
Análisis de regresión de Cox.....	79
Regresión de Cox: Definir variables categóricas.....	80
Regresión de Cox: Gráficos.....	80
Regresión de Cox: Guardar variables nuevas.....	81
Regresión de Cox: Opciones.....	81
Regresión de Cox: Definir evento para la variable de estado.....	82
Características adicionales del comando COXREG.....	82
Calcular covariable dependiente del tiempo.....	82
Para calcular una covariable dependiente del tiempo.....	83
Esquemas de codificación de variables categóricas.....	83
Desviación.....	83
Sencillo.....	84
Helmert.....	84
Diferencia.....	85
Polinómico.....	85
Repetido.....	86

Especial.....	86
Indicador.....	87
Estructuras de covarianza.....	87
Estadísticas Bayesianas.....	91
Inferencia de una muestra Bayesiana: Normal.....	92
Inferencia de una muestra Bayesiana: Binomial.....	96
Inferencia de una muestra Bayesiana: Poisson.....	98
Inferencia de muestra relacionada Bayesiana: Normal.....	100
Inferencia de muestra independiente Bayesiana.....	101
Inferencia Bayesiana sobre la correlación de Pearson.....	105
Inferencia Bayesiana sobre modelos de regresión lineal.....	108
ANOVA de un factor Bayesiana.....	113
Modelos de logaritmo lineal Bayesianos.....	116
Modelos de ANOVA de medidas repetidas de un factor Bayesiano.....	119
Regresión Ridge del kernel.....	121
Parámetros de kernel.....	123
Regresión Ridge de kernel: Opciones.....	123
Avisos.....	125
Marcas comerciales.....	126
Índice.....	129

Capítulo 1. Estadísticos avanzados

Las características siguientes de estadísticas avanzadas están incluidas en SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

Introducción a Estadísticas avanzadas

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

- GLM Multivariante amplía el modelo lineal general que proporciona GLM Univariante al permitir varias variables dependientes. Una extensión adicional, MLG Medidas repetidas, permite las mediciones repetidas de varias variables dependientes.
- Análisis de componentes de la varianza es una herramienta específica para descomponer la variabilidad 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 mediciones 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.

Análisis GLM multivariante

El procedimiento GLM 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. Con este procedimiento de modelo lineal general se puede 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 de 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 vectoriales). Si se especifica más de una variable dependiente, se proporciona el análisis multivariante 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, GLM 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 visualizar 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 están disponibles una matriz SCPC residual, que es una matriz cuadrada de las sumas de cuadrados y los productos vectoriales 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 le permite especificar una variable usada para aplicar a las observaciones distintas ponderaciones en un análisis de mínimos cuadrados ponderados (MCP), por ejemplo para compensar la distinta precisión de las mediciones.

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 detecta 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 se pueden emplear para evaluar las diferentes hipótesis. Tipo III es el valor predeterminado.

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 estándar 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. Dispersión versus nivel, residuos y perfil (interacción).

GLM Multivariante: Consideraciones sobre los datos

Datos. Las variables dependientes deben ser cuantitativas. Los factores son categóricos y pueden tener valores numéricos o valores de serie. 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 GLM Factorial General. Si ha medido las mismas variables dependientes en varias ocasiones para cada sujeto, utilice GLM Medidas repetidas.

Obtención de tablas de GLM Multivariante

1. En los menús seleccione:

Analizar > Modelo lineal general > Multivariante...

2. Seleccione al menos dos variables dependientes.

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

Modelo MLG 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 de suma de cuadrados más utilizado es el de 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.

Construcción de términos y Términos personalizados

Construir términos

Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados

Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

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 predeterminado.

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 de suma de cuadrados de Tipo II 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 predeterminado. Este método calcula las sumas de cuadrados de un efecto de diseño como las sumas de cuadrados, corregidas respecto a cualquier otro efecto que no contenga el efecto, y ortogonales a cualquier efecto (si existe) que contenga el efecto. 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 de suma de cuadrados de Tipo III 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 de suma de cuadrados de Tipo IV 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

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 $\mathbf{LBM} = \mathbf{0}$, donde \mathbf{L} es la matriz de coeficientes de contraste, \mathbf{M} es la matriz de identidad (que tiene una dimensión igual al número de variables dependientes) y \mathbf{B} es el vector de parámetros. Cuando se especifica un contraste, se crea una matriz \mathbf{L} de modo que las columnas correspondientes al factor coincidan con el contraste. El resto de las columnas se corrigen para que la matriz \mathbf{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 multivariantes 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.

Simples. 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).

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

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.

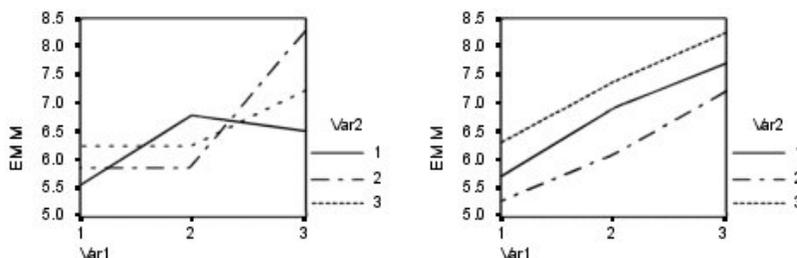


Figura 1. Gráfico no paralelo (izquierda) y gráfico paralelo (derecha)

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

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 predeterminada. 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 probar si la media en cualquier nivel del factor es menor que la de la categoría de control, seleccione **< Control**. Asimismo, para probar 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 parejas Games-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 de la muestra 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 característica. 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 de significación observado 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.

Medias marginales estimadas GLM

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.

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.

Comparar efectos principales simples

El valor se habilita siempre que la lista de destinos contenga uno o más efectos de producto o interacción (por ejemplo, A*B, A*B*C). El valor admite la especificación de comparaciones entre los efectos principales simples, que son los efectos principales anidados dentro de los niveles de otros factores.

Ajuste de intervalo de confianza

Seleccione un ajuste de diferencia menor significativa (DMS), Bonferroni o Sidak para los intervalos de confianza y la significación. Este elemento solo estará disponible si se selecciona **Comparar los efectos principales y/o Comparar efectos principales simples**.

Especificación de medias marginales estimadas

1. Seleccione en los menús uno de los procedimientos disponibles en > **Analizar** > **Modelo lineal general**.
2. En el cuadro de diálogo principal, pulse **Medias ME**.

Guardar MLG

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.

- *Sin estandarizar.* Valor predicho por el modelo para la variable dependiente.
- *Ponderado.* Los valores pronosticados no estandarizados ponderados. Sólo están disponibles si se seleccionó previamente una variable de ponderación MCP.
- *Error estándar.* Estimación de la desviación estándar 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.* Los valores de influencia no centrados. La influencia relativa de una observación en el ajuste del modelo.

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

- *Sin estandarizar.* Diferencia entre un valor observado y el valor predicho por el modelo.
- *Ponderado.* Los residuos no estandarizados ponderados. Sólo están disponibles si se seleccionó previamente una variable de ponderación MCP.
- *Estandarizado.* El residuo dividido por una estimación de su error estándar. Los residuos estandarizados, que son conocidos también como los residuos de Pearson o residuos estandarizados, tienen una media de 0 y una desviación estándar de 1.
- *Estudentizado.* El residuo dividido por una estimación de su desviación estándar que varía de un caso a otro, dependiendo de la distancia de los valores de cada caso en las variables independientes de las medias de las variables independientes. A veces se hace referencia a ellos como residuos estudentizados internamente.
- *Suprimido.* 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 predicho 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 IBM SPSS Statistics. Asimismo, para cada variable dependiente habrá una fila de estimaciones de los parámetros, una fila de errores estándar de las 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. Cuando se selecciona la estadística coherente de heterocedasticidad (sólo disponible para los modelos univariante), se calcula la matriz varianza-covarianza utilizando un estimador robusto, la fila de errores estándar muestra los errores estándar robustos, y los valores de significación reflejan los

errores robustos. Si lo desea, puede usar este archivo matricial en otros procedimientos que lean archivos matriciales.

MLG Multivariante: Opciones

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

Representación. Seleccione **Estadísticos descriptivos** para generar medias observadas, desviaciones estándar 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, errores estándar, pruebas *t*, 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 intra-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. **Funciones estimables generales** 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 o funciones estimables generales.

Visualización

Estadísticos descriptivos

Genera medias observadas, desviaciones estándar y recuentos para todas las variables dependientes en todas las casillas.

Estimaciones del tamaño del efecto

Proporciona un valor de eta-cuadrado parcial para cada efecto y cada estimación de parámetro. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor.

Potencia observada

Obtiene la potencia de la prueba cuando la hipótesis alternativa se establece basándose en el valor observado.

Estimaciones de los parámetros

Genera las estimaciones de parámetro, errores estándar, pruebas *t*, intervalos de confianza y la potencia observada para cada prueba.

Matrices SSCP

Muestra las matrices SSCP de error e hipótesis.

Matriz SSCP residual

Muestra la matriz SSCP residual de hipótesis y error.

Matriz de transformación

Muestra la prueba de esfericidad de Bartlett de la matriz de covarianza residual.

Pruebas de homogeneidad

Genera la prueba Levene de la homogeneidad de varianza para cada variable dependiente entre todas las combinaciones de nivel de los factores inter-sujetos, solo 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 intra-sujetos.

Gráfico de dispersión frente a gráfico de niveles

Es útil para comprobar los supuestos sobre los datos para investigar el supuesto de varianza igual. Estos elementos no estarán activado si no hay factores.

Gráfico de residuos

Genera un gráfico de residuos observado-por-predicho-por-estandarizado para cada variable dependiente. El gráfico es útil para investigar el supuesto de una varianza igual.

Falta de ajuste

Compruebe si la relación entre la variable dependiente y las variables independientes se pueden describir de forma adecuada con el modelo.

Funciones estimables generales

Le permite construir contrastes de hipótesis personalizados basados en la función o funciones estimables generales. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función o funciones estimables generales.

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. Al especificar un nivel de significación, el nivel asociado de los intervalos de confianza se muestra en el diálogo.

Características adicionales del mandato GLM

Estas características se pueden aplicar a los análisis univariados, multivariados o de medidas repetidas. La sintaxis de comandos también le permite:

- 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 del usuario (utilizando el subcomando MISSING).
- Especificar criterios EPS (utilizando el subcomando CRITERIA).
- Construir una matriz **L**, una matriz **M** o una matriz **K** (utilizando los subcomandos LMATRIX, MMATRIX o KMATRIX).
- Para contrastes de desviación o simples, especifique una categoría de referencia intermedia (utilizando el subcomando CONTRAST).
- Especificar métricas para contrastes polinómicos (utilizando el subcomando CONTRAST).
- Especificar términos de error para comparaciones post hoc (utilizando el subcomando POSTHOC).
- Calcular medias marginales estimadas para cualquier factor o interacción de factores entre los factores de la lista de factores (utilizando el subcomando EMMEANS).
- Especificar nombres para las variables temporales (utilizando el subcomando SAVE).
- Construir un archivo de datos de matriz de correlaciones (utilizando el subcomando OUTFILE).
- Construir un archivo de datos de matriz que contenga estadísticos de la tabla de ANOVA intra-sujetos (utilizando el subcomando OUTFILE).
- Guardar la matriz de diseño en un archivo de datos nuevo (utilizando el subcomando OUTFILE).

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

MLG Medidas repetidas

El procedimiento MLG Medidas repetidas proporciona un análisis de varianza cuando se toma la misma medición 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 mediciones 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 de 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 vectoriales). 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 visualizar 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 están disponibles una matriz SCPC residual, que es una matriz cuadrada de las sumas de cuadrados y los productos vectoriales 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 le permite especificar una variable usada para aplicar a las observaciones distintas ponderaciones en un análisis de mínimos cuadrados ponderados (MCP), por ejemplo para compensar la distinta precisión de las mediciones.

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 se pueden emplear para evaluar las diferentes hipótesis. Tipo III es el valor predeterminado.

Estadísticos. Las pruebas de rango post hoc y las comparaciones múltiples (para 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 estándar 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 y la prueba de esfericidad de Mauchly.

Diagramas. Dispersión versus nivel, residuos y perfil (interacción).

MLG Medidas repetidas: Consideraciones sobre los datos

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 mediciones tomadas a los sujetos. El conjunto tiene una variable para cada repetición de la medición 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 mediciones del peso en días diferentes. Si las mediciones 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 mediciones 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 mediciones 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 GLM 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 mediciones 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 mediciones repetidas para cada sujeto, utilice MLG Univariante o MLG Multivariante. Si sólo existen dos mediciones para cada sujeto (por ejemplo, antes del test y después del test) y no hay factores inter-sujetos, puede utilizar el procedimiento Prueba T para muestras emparejadas.

Obtención de MLG Medidas repetidas

1. En los menús seleccione:

Analizar > Modelo lineal general > Medidas repetidas...

2. Escriba un nombre para el factor intra-sujetos y su número de niveles.

3. Pulse **Add**.

4. Repita estos pasos para cada factor intra-sujetos.

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

5. Escriba el nombre de la medida.

6. Pulse **Add**.

Después de definir todos los factores y las medidas:

7. Pulse en **Definir**.

8. 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 mediciones del mismo atributo. Este cuadro de diálogo permite definir uno o varios factores intra-sujetos para utilizarlos en MLG Medidas repetidas. 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

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 de suma de cuadrados más utilizado es el de Tipo III.

Construcción de términos y Términos personalizados

Construir términos

Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados

Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

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 predeterminado.

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 de suma de cuadrados de Tipo II 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 predeterminado. Este método calcula las sumas de cuadrados de un efecto de diseño como las sumas de cuadrados, corregidas respecto a cualquier otro efecto que no contenga el efecto, y ortogonales a cualquier efecto (si existe) que contenga el efecto. 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 de suma de cuadrados de Tipo III 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 de suma de cuadrados de Tipo IV 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

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 $\mathbf{LBM} = 0$, donde \mathbf{L} es la matriz de coeficientes de contraste, \mathbf{B} es el vector de parámetros y \mathbf{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 predeterminado) para los factores intra-sujetos, la matriz \mathbf{M} será (0,5 0,5 0,5 0,5)⁴. Cuando se especifica un contraste, se crea una matriz \mathbf{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 \mathbf{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.

Simples. 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).

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

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.

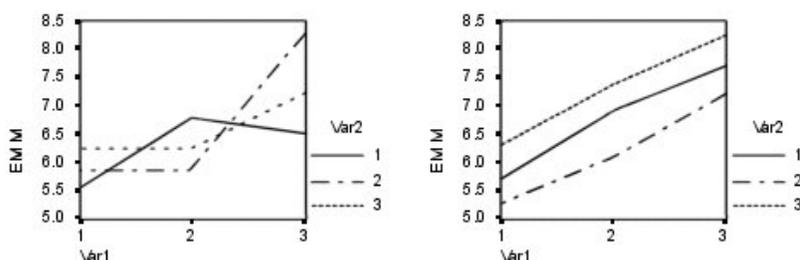


Figura 2. Gráfico no paralelo (izquierda) y gráfico paralelo (derecha)

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

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 predeterminada. 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 probar si la media en cualquier nivel del factor es menor que la de la categoría de control, seleccione **< Control**. Asimismo, para probar 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 parejas Games-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 de la muestra 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 característica. 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 de significación observado 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.

Medias marginales estimadas GLM

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.

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.

Comparar efectos principales simples

El valor se habilita siempre que la lista de destinos contenga uno o más efectos de producto o interacción (por ejemplo, $A*B$, $A*B*C$). El valor admite la especificación de comparaciones entre los efectos principales simples, que son los efectos principales anidados dentro de los niveles de otros factores.

Ajuste de intervalo de confianza

Seleccione un ajuste de diferencia menor significativa (DMS), Bonferroni o Sidak para los intervalos de confianza y la significación. Este elemento solo estará disponible si se selecciona **Comparar los efectos principales** y/o **Comparar efectos principales simples**.

Especificación de medias marginales estimadas

1. Seleccione en los menús uno de los procedimientos disponibles en > **Analizar** > **Modelo lineal general**.
2. En el cuadro de diálogo principal, pulse **Medias ME**.

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, debe guardar el archivo de datos actual.

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

- *Sin estandarizar.* Valor predicho por el modelo para la variable dependiente.
- *Error estándar.* Estimación de la desviación estándar 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.* Los valores de influencia no centrados. La influencia relativa de una observación en el ajuste del modelo.

Residuos. Un residuo no estandarizado es el valor real de la variable dependiente menos el valor predicho por el modelo. También se encuentran disponibles residuos eliminados, estudentizados y estandarizados.

- *Sin estandarizar.* Diferencia entre un valor observado y el valor predicho por el modelo.
- *Estandarizado.* El residuo dividido por una estimación de su error estándar. Los residuos estandarizados, que son conocidos también como los residuos de Pearson o residuos estandarizados, tienen una media de 0 y una desviación estándar de 1.
- *Estudentizado.* El residuo dividido por una estimación de su desviación estándar que varía de un caso a otro, dependiendo de la distancia de los valores de cada caso en las variables independientes de las medias de las variables independientes. A veces se hace referencia a ellos como residuos estudentizados internamente.
- *Suprimido.* 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 predicho 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 guardan 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. .

GLM Medidas repetidas: Opciones

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

Visualización

Estadísticos descriptivos

Genera medias observadas, desviaciones estándar y recuentos para todas las variables dependientes en todas las casillas.

Estimaciones del tamaño del efecto

Proporciona un valor de eta-cuadrado parcial para cada efecto y cada estimación de parámetro. El estadístico eta cuadrado describe la proporción de variabilidad total atribuible a un factor.

Potencia observada

Obtiene la potencia de la prueba cuando la hipótesis alternativa se establece basándose en el valor observado.

Estimaciones de los parámetros

Genera las estimaciones de parámetro, errores estándar, pruebas t , intervalos de confianza y la potencia observada para cada prueba.

Matrices SSCP

Muestra las matrices SSCP de error e hipótesis.

Matriz SSCP residual

Muestra la matriz SSCP residual.

Matriz de transformación

Muestra la prueba de esfericidad de Bartlett de la matriz de covarianza residual.

Pruebas de homogeneidad

Genera la prueba Levene de la homogeneidad de varianza para cada variable dependiente entre todas las combinaciones de nivel de los factores inter-sujetos, solo 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 intra-sujetos.

Diagrama de dispersión por nivel

Es útil para comprobar supuestos sobre los datos. Esta opción está inhabilitada si no hay ningún factor.

Gráfico de residuos

Genera un gráfico de residuos observado-por-predicho-por-estandarizado para cada variable dependiente. Estos gráficos son útiles para investigar el supuesto de varianzas iguales. Esta opción está inhabilitada si no hay ningún factor.

Falta de ajuste

Comprueba si la relación entre la variable dependiente y las variables independientes se puede describir de forma adecuada por el modelo.

Funciones estimables generales

Le permite construir contrastes de hipótesis personalizados basados en la función o funciones estimables generales. Las filas en las matrices de coeficientes de contraste son combinaciones lineales de la función o funciones estimables generales.

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. Al especificar un nivel de significación, el nivel asociado de los intervalos de confianza se muestra en el diálogo.

Características adicionales del mandato GLM

Estas características se pueden aplicar a los análisis univariados, multivariados o de medidas repetidas. La sintaxis de comandos también le permite:

- 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 varios contrastes (utilizando el subcomando CONTRAST).
- Incluir los valores perdidos del 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).
- Para contrastes de desviación o simples, especifique una categoría de referencia intermedia (utilizando el subcomando CONTRAST).
- Especificar métricas para contrastes polinómicos (utilizando el subcomando CONTRAST).
- Especificar términos de error para comparaciones post hoc (utilizando el subcomando POSTHOC).
- Calcular medias marginales estimadas para cualquier factor o interacción de factores entre los factores de la lista de factores (utilizando el subcomando EMMEANS).
- Especificar nombres para variables temporales (utilizando el subcomando SAVE).
- Construir un archivo de datos de matriz de correlaciones (utilizando el subcomando OUTFILE).
- Construir un archivo de datos de matriz 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).

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

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 predeterminados 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 el historial 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 mediciones.

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.

Componentes de la varianza: Consideraciones sobre los datos

Datos. La variable dependiente es cuantitativa. Los factores son categóricos. 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

1. En los menús seleccione:

Analizar > Modelo lineal general > Componentes de la varianza...

2. Seleccione una variable dependiente.

3. 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

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.

Para las covariables y los factores seleccionados:

Interacción

Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales

Crea un término de efectos principales para cada variable seleccionada.

Todos los de 2 factores

Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todos los de 3 factores

Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todos los de 4 factores

Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todos los de 5 factores

Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

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

Construcción de términos y Términos personalizados

Construir términos

Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados

Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

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 estándar 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 predeterminado 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.

Representación. 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 predeterminado.

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 de suma de cuadrados de Tipo I 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. El valor predeterminado. Este método calcula las sumas de cuadrados de un efecto de 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 de suma de cuadrados de Tipo III 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

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 guardan 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.

Características adicionales del comando VARCOMP

La sintaxis de comandos también le permite:

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

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

Modelos lineales mixtos

El procedimiento Modelos lineales mixtos amplía 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.

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 estándar 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 predeterminado.

Consideraciones sobre datos de modelos mixtos lineales

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

1. En los menús seleccione:

Analizar > Modelos mixtos > Lineal...

2. Si lo desea, puede seleccionar una o más variables de sujetos.

3. Si lo desea, puede seleccionar una o más variables repetidas. Si se han definido variables repetidas, seleccione un **Tipo de covarianza repetido** en la lista desplegable.

4. Si lo desea, puede seleccionar una o más variables de medidas Kronecker.

5. Si lo desea, puede seleccionar una estructura de covarianza residual.

6. Pulse **Continuar**.

7. Seleccione una variable dependiente.

8. Seleccione al menos un factor o covariable.

9. Pulse **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: Sujetos y Repetidos

Este diálogo le permite seleccionar variables que definen sujetos, observaciones repetidas, medidas Kronecker y para elegir una estructura de covarianza para los residuos.

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 sujetos resulta especialmente importante cuando se producen mediciones repetidas por sujeto y quiere modelar la correlación entre estas observaciones. Por ejemplo, cabría esperar que las lecturas de la tensión arterial de un único paciente durante visitas consecutivas al médico estén correlacionadas.

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 utilizan para definir sujetos para la estructura de covarianzas de residuos. Puede usar todas o algunas de las variables que definen los sujetos para la estructura de la covarianza de los efectos aleatorios.

Repetido

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 repetido

Esto especifica la estructura de covarianzas para los residuos. Las estructuras disponibles son las siguientes:

- Dependencia Ante: Primer orden
- AR(1)
- Producto directo AR1 (UN_AR1)
- Producto directo sin estructura (UN_UN)
- Simetría compuesta de producto directo(UN_CS)
- 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
- Espacial: potencia
- Espacial: exponencial
- Espacial: gaussiano
- Espacial: Lineal
- Espacial: log-lineal
- Espacial: esférico

Medidas Kronecker

Seleccione las variables que especifican la estructura del sujeto para las mediciones de covarianza Kronecker y determine cómo se correlacionan los errores de medición. El campo solo está disponible cuando se selecciona uno de los siguientes **Tipo de covarianza repetido**:

- Producto directo AR1 (UN_AR1)
- Producto directo sin estructura (UN_UN)
- Simetría compuesta de producto directo(UN_CS)

Coordenadas de covarianza espaciales

Las variables de esta lista especifican las coordenadas de las observaciones repetidas, cuando se selecciona uno de los tipos de covarianza espaciales para el tipo de covarianza repetido.

Consulte el tema “Estructuras de covarianza” en la página 87 para obtener más información.

Efectos fijos de los Modelos lineales mixtos

Efectos fijos. No existe un modelo predeterminado, 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.

Generar 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 valor predeterminado.

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 bidimensionales posibles de las variables seleccionadas.

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

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

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

Crear términos anidados

En este procedimiento puede generar términos anidados para su modelo. Los términos anidados son útiles para modelar el efecto de un factor o covariable cuyos valores no interactúen 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. Los términos anidados tienen las siguientes restricciones:

- Todos los factores incluidos en una interacción deben ser exclusivos. Por consiguiente, si A es un factor, no es válido especificar A^*A .
- Todos los factores incluidos en un efecto anidado deben ser exclusivos. 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 predeterminado.

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 predeterminado. Este método calcula las sumas de cuadrados de un efecto de 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 de suma de cuadrados de Tipo III 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

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
- Componentes de la varianza

Consulte el tema [“Estructuras de covarianza”](#) en la [página 87](#) para obtener más información.

Efectos aleatorios. No existe un modelo predeterminado, 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 generado el primer modelo, pulse en **Siguiente** para generar 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.

Mostrar predicciones de parámetros para este conjunto de efectos aleatorios. Especifica que se visualicen las estimaciones de los parámetros de efectos aleatorios.

Estimación de los Modelos lineales mixtos

Método

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

Grados de libertad

Proporciona opciones para definir los grados de libertad para todas las pruebas.

Método de residuo

El método de residuo tiene unos grados de libertad fijos para todas las pruebas. Es útil si el tamaño de su muestra es suficientemente grande, si los datos están equilibrados o si el modelo utiliza un tipo de covarianza más simple (por ejemplo, identidad escalada o diagonal).

Aproximación Satterthwaite

El método Satterthwaite tiene grados de libertad de un campo en todas las pruebas. Es útil si el tamaño de la muestra es pequeño, si los datos no están equilibrados o si el modelo utiliza un tipo de covarianza complicado (por ejemplo, sin estructura).

Aproximación Kenward-Roger

El método Kenward-Roger ofrece un estimador de muestras pequeño más preciso para la varianza-covarianza de los parámetros de efectos fijo y los grados de libertad de denominador aproximados en pruebas t y pruebas F . El método introduce un factor de escala para el estadístico F y lo estima y los grados de libertad del denominador utilizando una expansión de la serie de Taylor para la estructura aleatoria estimada dentro de los datos.

Nota: El método Kenward-Roger se utiliza en la covarianza basada en el modelo (en lugar de la covarianza robusta). Cuando se selecciona el método Kenward-Roger y la covarianza robusta, el método Kenward-Roger se aplica a la covarianza basada en modelo, y se muestra el siguiente aviso: "Puesto que se ha seleccionado el método Kenward-Roger, el método de covarianza sólida se ha modificado al método de covarianza basada en modelo."

Iteraciones

Están disponibles las opciones siguientes:

Máximo de iteraciones

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. Establézcalo en un entero positivo.

Imprimir historial de iteraciones para cada n paso(s)

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^a (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 el cambio relativo en la función de log-verosimilitud es menor que el valor especificado, que no debe ser 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 de puntuación

Solicita utilizar el algoritmo de puntuación de Fisher hasta el número de iteración n . Especifique un entero no negativo.

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

Estadísticos de resumen. Genera tablas correspondientes a:

- **Estadísticos descriptivos.** Muestra los tamaños de las muestras, medias y desviaciones estándar 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 parámetros de los efectos fijos.** Muestra las estimaciones de los parámetros de los efectos fijos y los errores estándar aproximados correspondientes.
- **Contrastes sobre parámetros de covarianza.** Muestra los errores estándar 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 predeterminado es 95.

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 estándar 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 contruidos 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

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

Residuos y valores pronosticados. Guarda las variables relacionadas con el valor ajustado por el modelo.

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

Características adicionales del comando MIXED

La sintaxis de comandos también le permite:

- Especificar contrastes de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando TEST).
- Incluir los valores perdidos del 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).

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

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 no 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 cuáles 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.

Modelos lineales generalizados: Consideraciones sobre los datos

Datos. La respuesta puede ser de escala, de recuentos, binaria o eventos en ensayos. Se supone que los factores son categóricos. Las covariables, la ponderación 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

En los menús seleccione:

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

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

- **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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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; puede establecerlo en un valor fijo o permitir que se calcule mediante 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 se puede considerar como el número de ocurrencias de un suceso 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. Están disponibles las opciones siguientes:

- **Identidad.** $f(x)=x$. No se transforma la variable dependiente. Este enlace se puede utilizar con cualquier distribución.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Resulta apropiada únicamente para la distribución binomial.

- **Cauchit acumulado.** $f(x) = \tan[\pi(x - 0.5)]$, aplicado a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Log-log complementario acumulado.** $f(x) = \ln[-\ln(1-x)]$, aplicado 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.** $f(x) = -\ln[-\ln(x)]$, aplicado 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 Φ^{-1} es la función de distribución acumulada normal estándar inversa. Es apropiada únicamente para la distribución multinomial.
- **Registro.** $f(x) = \log(x)$. Este enlace se puede utilizar con cualquier distribución.
- **Complemento de registro.** $f(x) = \log(1-x)$. Resulta apropiada únicamente para la distribución binomial.
- **Logit.** $f(x) = \log(x / (1-x))$. Resulta 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.** $f(x) = -\log(-\log(x))$. Resulta apropiada únicamente para la distribución binomial.
- **Potencia de probabilidades.** $f(x) = \{[x/(1-x)]^\alpha - 1\} / \alpha$, si $\alpha \neq 0$. $f(x) = \log(x)$, si $\alpha = 0$. α es la especificación del número requerido y debe ser un número real. Resulta apropiada únicamente para la distribución binomial.
- **Probit.** $f(x) = \Phi^{-1}(x)$, donde Φ^{-1} es la función de distribución acumulada normal estándar inversa. Resulta apropiada únicamente para la distribución binomial.
- **Potencia.** $f(x) = x^\alpha$, si $\alpha \neq 0$. $f(x) = \log(x)$, si es $\alpha = 0$. α es la especificación del número requerido y debe ser un número real. Este enlace se puede utilizar con cualquier distribución.

Modelos lineales generalizados: 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).

Ponderación 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 ponderación 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 de ponderación de escala es menor o igual que 0 o que son perdidos no se utilizan en el análisis.

Modelos lineales generalizados: Categoría de referencia

Para la respuesta binaria, puede seleccionar la categoría de referencia para 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:

- De forma predeterminada, 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 categoría de referencia, entonces las probabilidades guardadas por el modelo calculan la posibilidad de que un caso determinado 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

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 desviación 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 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 Poisson o una binomial negativa con un enlace de registro si el registro natural de la experiencia del conductor se incluye como un término de desplazamiento.

Otras combinaciones de los tipos de distribución y enlace requerirán otras transformaciones de la variable de desplazamiento.

Modelos lineales generalizados: Opciones

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

Valores perdidos del usuario. Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores perdidos del 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 exclusivo encontrado define la última categoría.

Modelos lineales generalizados: Modelo

Especificar efectos del modelo

El modelo predeterminado 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.

Todos los de 2 factores

Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todos los de 3 factores

Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todos los de 4 factores

Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todos los de 5 factores

Crea todas las interacciones quintuples posibles de las variables seleccionadas.

Términos anidados

En este procedimiento puede generar términos anidados para su modelo. Los términos anidados son útiles para modelar el efecto de un factor o covariable cuyos valores no interactúen 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: Los términos anidados tienen las siguientes restricciones:

- Todos los factores incluidos en una interacción deben ser exclusivos. Por consiguiente, si *A* es un factor, no es válido especificar $A*A$.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos. 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 $A(X)$.

Modelos lineales generalizados: 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 Puntuación 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 desviación 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. También puede especificar los valores iniciales para 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 de Huber/White/sandwich) es un estimador "corregido" basado en el modelo que proporciona una estimación coherente de la covarianza, incluso cuando se ha especificado incorrectamente la varianza y las funciones de enlace.

Iteraciones. Están disponibles las opciones siguientes:

- **Número máximo de iteraciones.** 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. Establézcalo en un 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. Se encuentran disponibles las siguientes opciones:

- **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 puede 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

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: *RowType_*, *VarName_*, *P1*, *P2*..., donde *RowType_* y *VarName_* son variables de serie, 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*, etc. 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 de archivos el orden especificado al crear la segmentación de archivos, seguidas de *TipoFila_*, *NombreVar_*, *P1*, *P2*, etc., 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. para obtener más información.

Nota: los nombres de las variables *P1*, *P2*, etc., no son necesarios. El procedimiento aceptará cualquier nombre de variable válido para los parámetros, ya que la correlación 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

Efectos del modelo. Están disponibles las opciones siguientes:

- **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 desviación y la desviación escalada, chi-cuadrado de Pearson y chi-cuadrado de Pearson escalado, log-verosimilitud, criterio de información de Akaike (AIC), AIC corregido para muestras finitas (AICC), criterio de información bayesiano (BIC) y AIC consistente (CAIC).
- **Estadísticos de resumen del modelo.** Muestra contraste de ajuste del modelo, incluidos los estadísticos de la razón de la verosimilitud para la prueba 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 predeterminados y para las medias marginales estimadas, si se solicitaron en el separador 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 el 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^{a} iteraciones a partir de la iteración 0^{a} (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 de parámetro de escala o parámetro auxiliar binomial negativa.** Muestra los estadísticos de prueba de multiplicador de Lagrange para evaluar la validez de un parámetro de escala que se calcula utilizando la desviación 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, se contrasta el parámetro auxiliar fijo.

Modelos lineales generalizados: Medias marginales estimadas

Esta pestaña permite ver medias marginales estimadas para los niveles de factores y las interacciones de 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 (a veces también se denominan contrastes de Helmert inversos). 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.

Escalas. 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.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de 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.* Éste 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

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 exclusivos.

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 de intervalo de confianza para 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 está 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 predicho de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.

Error estándar estimado del valor pronosticado del predictor lineal

Cuando la distribución de respuesta es multinomial, el procedimiento guarda el error estándar 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 $(N-1)/N$.

Residual

Diferencia entre un valor observado y el valor predicho 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 estandarizado

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 desviación

La raíz cuadrada de la contribución de un caso al estadístico de desviación, con el signo del residuo bruto.

Residuo de desviación estandarizado

El residuo de desviación 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 desviación, con el signo del residuo bruto.

Modelos lineales generalizados: 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 estándar, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **Variáveis de segmentación.** Si se han utilizado, todas las variables que definan segmentaciones.
- **RowType_.** Toma valores (y etiquetas de valor) *COV* (covarianzas), *CORR* (correlaciones), *EST* (estimaciones de parámetros), *SE* (errores estándar), *SIG* (niveles de significancia) y *DF* (grados de libertad de diseño de muestreo). 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 valores *P1*, *P2*... correspondientes a una lista ordenada de todos los parámetros de modelo estimados (excepto los parámetros de escala o de binomial negativa), para los tipos de fila *COV* o *CORR*, con etiquetas de valor correspondientes a las series de parámetros que se muestran en la tabla de estimaciones de 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 estándar, 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 estándar; 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 estándar; 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 estándar, 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. si desea más información.

Características 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.

Ecuaciones de estimación generalizadas

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

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.

Ecuaciones de estimación generalizadas: Consideraciones sobre los datos

Datos. La respuesta puede ser de escala, de recuentos, binaria o eventos en ensayos. Se supone que los factores son categóricos. Las covariables, la ponderación de escala y el desplazamiento se suponen que son de escala. Las variables utilizadas para definir los sujetos o las mediciones 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 intra-sujetos e independientes inter-sujetos. La matriz de correlaciones que representa las dependencias intra-sujetos se estima como parte del modelo.

Obtención de ecuaciones de estimación generalizadas

En los menús seleccione:

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

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

La combinación de valores de las variables especificadas debe definir de forma exclusiva 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 exclusivos 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.

2. En la pestaña Tipo de modelo, especifique una función de enlace y distribución.
3. En la pestaña Respuesta, seleccione una variable dependiente.
4. En la pestaña Predictores, seleccione factores y covariables que utilizará para pronosticar la variable dependiente.
5. En la pestaña Modelo, especifique los efectos del modelo utilizando las covariables y factures 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 mediciones dentro de los sujetos. Por tanto, la combinación de las variables intra-sujetos y de los sujetos define de forma exclusiva cada medición. 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 mediciones 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 mediciones 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 mediciones y, por tanto, por la combinación de los valores de las variables intra-sujetos. Puede especificar una de las siguientes estructuras:

- **Independiente.** Las mediciones repetidas no están correlacionadas.
- **AR(1).** Las mediciones repetidas tienen una relación autorregresiva de primer orden. La correlación entre dos elementos cualesquiera es igual a ρ para elementos adyacentes, ρ^2 para elementos separados por un tercero, y así sucesivamente y está restringido de modo 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 mediciones 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. Por ejemplo, si proporciona pruebas estandarizadas a los alumnos cada año desde el 3º hasta el 7º grado. Esta estructura supone que las puntuaciones de los grados 3º y 4º, 4º y 5º, 5º y 6º, y 6º y 7º tendrán la misma correlación; 3º y 5º, 4º y 6º, y 5º y 7º tendrán la misma correlación; 3º y 6º y 4º y 7º tendrán la misma correlación. Las medidas con una separación mayor que m se supone que 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.

De forma predeterminada, 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 réplica 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

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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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. Están disponibles las opciones siguientes:

- **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; puede establecerlo en un valor fijo o permitir que se calcule mediante 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 se puede considerar como el número de ocurrencias de un suceso 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. Están disponibles las opciones siguientes:

- **Identidad.** $f(x)=x$. No se transforma la variable dependiente. Este enlace se puede utilizar con cualquier distribución.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Resulta apropiada únicamente para la distribución binomial.

- **Cauchit acumulado.** $f(x) = \tan[\pi(x - 0.5)]$, aplicado a la probabilidad acumulada de cada categoría de la respuesta. Es apropiada únicamente para la distribución multinomial.
- **Log-log complementario acumulado.** $f(x) = \ln[-\ln(1-x)]$, aplicado 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.** $f(x) = -\ln[-\ln(x)]$, aplicado 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 Φ^{-1} es la función de distribución acumulada normal estándar inversa. Es apropiada únicamente para la distribución multinomial.
- **Registro.** $f(x) = \log(x)$. Este enlace se puede utilizar con cualquier distribución.
- **Complemento de registro.** $f(x) = \log(1-x)$. Resulta apropiada únicamente para la distribución binomial.
- **Logit.** $f(x) = \log(x / (1-x))$. Resulta 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.** $f(x) = -\log(-\log(x))$. Resulta apropiada únicamente para la distribución binomial.
- **Potencia de probabilidades.** $f(x) = \{[x/(1-x)]^\alpha - 1\} / \alpha$, si $\alpha \neq 0$. $f(x) = \log(x)$, si $\alpha = 0$. α es la especificación del número requerido y debe ser un número real. Resulta apropiada únicamente para la distribución binomial.
- **Probit.** $f(x) = \Phi^{-1}(x)$, donde Φ^{-1} es la función de distribución acumulada normal estándar inversa. Resulta apropiada únicamente para la distribución binomial.
- **Potencia.** $f(x) = x^\alpha$, si $\alpha \neq 0$. $f(x) = \log(x)$, si es $\alpha = 0$. α es la especificación del número requerido y debe ser un número real. Este enlace se puede utilizar con cualquier distribución.

Respuesta de las ecuaciones de estimación generalizadas

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).

Ponderación de escala. El parámetro de escala es un parámetro de 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 ponderación de escala, el parámetro de escala, que está relacionado con la varianza de la respuesta, se divide por ella para cada observación. Los casos cuyo valor de ponderación 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

Para la respuesta binaria, puede seleccionar la categoría de referencia para 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:

- De forma predeterminada, 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 categoría de referencia, entonces las probabilidades guardadas por el modelo calculan la posibilidad de que un caso determinado 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

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 desviación 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 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 Poisson o una binomial negativa con un enlace de registro si el registro natural de la experiencia del conductor se incluye como un término de desplazamiento.

Otras combinaciones de los tipos de distribución y enlace requerirán otras transformaciones de la variable de desplazamiento.

Ecuaciones de estimación generalizadas: Opciones

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

Valores perdidos del usuario. Los factores deben tener valores válidos para el caso para que se incluyan en el análisis. Estos controles permiten decidir si los valores perdidos del 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 exclusivo encontrado define la última categoría.

Ecuaciones de estimación generalizadas: Modelo

Especificar efectos del modelo. El modelo predeterminado 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 bidimensionales posibles de las variables seleccionadas.

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

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

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

Términos anidados

En este procedimiento puede generar términos anidados para su modelo. Los términos anidados son útiles para modelar el efecto de un factor o covariable cuyos valores no interactúen 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. Los términos anidados tienen las siguientes restricciones:

- Todos los factores incluidos en una interacción deben ser exclusivos. Por consiguiente, si A es un factor, no es válido especificar $A*A$.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos. 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

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, Puntuación de Fisher o un método híbrido en el que las iteraciones de Puntuación de Fisher se realizan antes de cambiar al método de Newton-Raphson. Si se logra la convergencia durante la fase de Puntuación 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 desviación 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. También puede especificar los valores iniciales para 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. Están disponibles las opciones siguientes:

- **Número máximo de iteraciones.** 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. Establézcalo en un 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. Se encuentran disponibles las siguientes opciones:

- **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 puede 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.

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: *RowType_*, *VarName_*, *P1*, *P2*..., donde *RowType_* y *VarName_* son variables de serie, 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*, etc. 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 de archivos el orden especificado al crear la segmentación de archivos, seguidas de *TipoFila_*, *NombreVar_*, *P1*, *P2*, etc., 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. para obtener más información.

Nota: los nombres de las variables *P1*, *P2*, etc., no son necesarios. El procedimiento aceptará cualquier nombre de variable válido para los parámetros, ya que la correlación 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

Efectos del modelo. Están disponibles las opciones siguientes:

- **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 en 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 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 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 la prueba 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 predeterminados y para las medias marginales estimadas, si se solicitaron en el separador 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 el 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^{a} iteraciones a partir de la iteración 0^{a} (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

Esta pestaña permite ver medias marginales estimadas para los niveles de factores y las interacciones de 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 seleccionarse 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 (a veces también se denominan contrastes de Helmert inversos). 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.

Escalas. 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.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de 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.* Éste 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

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 exclusivos. para obtener más información.

- **Valor predicho 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 de intervalo de confianza para la media de la respuesta.** Guarda el límite superior de intervalo de confianza para 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 está disponible para otras distribuciones de la respuesta.
- **Valor predicho 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 predicho de cada categoría de la respuesta, excepto la última, hasta el número de categorías que se ha especificado que se guarden.

- **Error estándar estimado del valor predicho del predictor lineal.** Cuando la distribución de respuesta es multinomial, el procedimiento guarda el error estándar 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 predicho 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

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 estándar, 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 valores (y etiquetas de valor) *COV* (covarianzas), *CORR* (correlaciones), *EST* (estimaciones de parámetros), *SE* (errores estándar), *SIG* (niveles de significancia) y *DF* (grados de libertad de diseño de muestreo). 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 valores *P1*, *P2*... correspondientes a una lista ordenada de todos los parámetros de modelo estimados (excepto los parámetros de escala o de binomial negativa), para los tipos de fila *COV* o *CORR*, con etiquetas de valor correspondientes a las series de parámetros que se muestran en la tabla de estimaciones de 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 estándar, 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 estándar; 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 estándar; 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 estándar, 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. si desea más información.

Características 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.

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 se pueden correlacionar.

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.

ejemplos

El consejo escolar del distrito puede utilizar un modelo mixto lineal generalizado para determinar si un método de enseñanza experimental es eficaz para mejorar las calificaciones de matemáticas. Los estudiantes de la misma aula deben correlacionarse, ya que reciben la enseñanza del mismo profesor, y además las aulas de la misma escuela deben también correlacionarse, de modo que se puedan incluir efectos aleatorios en los niveles de la escuela y las clases para explicar los diversos orígenes de variabilidad.

Los investigadores médicos pueden utilizar un modelo mezclado lineal generalizado para determinar si un nuevo fármaco anticonvulsivo puede reducir el índice de ataques epilépticos de un paciente. Normalmente, se han correlacionado positivamente repetidas mediciones del mismo paciente, así que debería ser apropiado un modelo mixto con algunos efecto aleatorios. El campo objetivo, que es el número de ataques, recibe valores enteros positivos, de modo que es posible que sea apropiado un modelo mixto lineal generalizado con una distribución Poisson y un enlace de logaritmo.

Los ejecutivos de un proveedor de televisión por cable, teléfono y servicios de Internet puede utilizar un modelo mixto lineal generalizado para conocer más detalles sobre clientes potenciales. Ya que las posibles respuestas tienen niveles de medición nominales, el analista de la empresa utiliza un modelo mixto logit generalizado con una intersección aleatoria para capturar la correlación entre respuestas a las preguntas de uso de servicios entre los tipos de servicios (televisión, teléfono, Internet) dentro de las respuestas de un encuestado específico.

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.

Opciones de efectos

Sujetos

La combinación de valores de los campos categóricos especificados debe definir de forma exclusiva 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 exclusivos 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 mediciones 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 el diálogo Variables se utilizan para definir sujetos para la estructura de covarianzas residual y proporcionan la lista de campos posibles para definir sujetos para estructuras de covarianza de 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.

Opciones de covarianza

Definir grupos de covarianzas por

Los campos categóricos especificados aquí definen conjuntos independientes de parámetros de covarianza de efectos repetidos; uno para 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 repetido

Especifica la estructura de la covarianza para los residuos. Están disponibles distintas opciones de covarianza en función del **Tipo de covarianza repetido**. Las estructuras disponibles son:

- Autorregresiva de primer orden (AR1)
- Producto directo AR1(UN_AR1)
- Producto directo sin estructura (UN_UN)
- Simetría compuesta de producto directo(UN_CS)
- Simetría compuesta heterogénea (CSH)
- Autoregresiva heterogénea(ARH1)
- Media móvil autorregresiva (1,1) (ARMA11)
- Simetría compuesta
- Diagonal
- Identidad escalada
- Toeplitz
- Sin estructura
- Componentes de la varianza
- Espacial: Potencia
- Espacial: Exponencial
- Espacial: De Gauss
- Espacial: Lineal
- Espacial: Log lineal
- Espacial: Esférico

Medidas Kronecker

Seleccione las variables que especifican la estructura del sujeto para las mediciones de covarianza Kronecker y determine cómo se correlacionan los errores de medición. El campo solo está disponible cuando se selecciona uno de los siguientes **Tipo de covarianza repetido**:

- Producto directo AR1(UN_AR1)
- Producto directo sin estructura (UN_UN)
- Simetría compuesta de producto directo(UN_CS)

Coordenadas de covarianza espacial

En las variables en esta lista se especifican las coordenadas de las observaciones repetidas cuando uno de los tipos de covarianza espacial está seleccionado para el tipo de covarianza para Repetidas.

Consulte el tema “Estructuras de covarianza” en la página 87 para obtener más información.

Pseudo R² medidas

Pseudo-R² medidas y el coeficiente de correlación intra-clase se incluyen en la salida GLMM (cuando corresponda). Las medidas pseudo-R² se basan totalmente en las estimaciones finales y se generan después de que se haya completado la estimación. El coeficiente de determinación R² es un estadístico notificado comúnmente, porque representa la proporción de varianza explicada por un modelo lineal. El coeficiente de correlación intra-clase (ICC) es una estadística relacionada que cuantifica la proporción de varianza explicada por un factor de agrupación (aleatorio) en datos multinivel/jerárquicos.

Obtención de un modelo mixto lineal generalizado

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

En los menús seleccione:

Analizar > Modelos mixtos > Lineal generalizado...

1. Defina la estructura del sujeto de su conjunto de datos en la pestaña **Estructura de datos**. La pestaña proporciona opciones para especificar los efectos de modelo y el tipo de covarianza repetida.
2. 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.
3. Pulse **Opciones de generación** para especificar cualquier configuración de generación.
4. Pulse en **Opciones de modelo** para guardar puntuaciones en el conjunto de datos activo y exportar el modelo en un archivo externo.
5. Pulse en **Ejecutar** para ejecutar el procedimiento y crear los objetos Modelo.

Objetivo

Estos ajustes definen el objetivo, su distribución y su relación con los predictores mediante la función de enlace.

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

- **Utilice el número de ensayos como denominador.** Cuando la respuesta objetivo es un número de sucesos que se producen en un conjunto de ensayos, el campo objetivo contiene el número de sucesos y se puede seleccionar un campo 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 objetivo (sucesos) 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, entonces el número de ensayos puede especificarse mediante un valor fijo.

El número de ensayos debe ser superior o igual al número de sucesos de cada registro. Los eventos deben ser enteros no negativos y los ensayos deben ser enteros positivos.

- **Personalice la categoría de referencia.** Para un objetivo categórico, puede seleccionar la categoría de referencia. Esto puede afectar a determinados 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, de forma predeterminada el procedimiento realiza la última categoría (la de mayor valor) o 2, la categoría de referencia. En esta situación, las estimaciones de los parámetros deben interpretarse en relación con la probabilidad de la categoría 0 o 1 *relativo* a la probabilidad de la categoría 2. Si especifica una categoría personalizada y el destino tiene etiquetas definidas, puede establecer la categoría de referencia eligiendo un valor de la lista. Esto puede resultar cómodo cuando, al especificar un modelo, no se recuerda exactamente cómo se codificó un determinado campo.

Distribución de objetivos y relación (enlace) con el modelo lineal. Dados los valores de los predictores, el modelo espera que la distribución de valores del objetivo siga la forma especificada y que los valores de objetivo estén linealmente relacionados 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 resulta de utilidad cuando se puede predecir el objetivo mediante un modelo de regresión lineal o ANOVA.
- **Regresión gamma.** Especifica una distribución gamma con un enlace de registro, que se debe utilizar si el objetivo contiene todos los valores positivos y es asimétrico a valores mayores.
- **Loglineal.** Especifica una distribución de Poisson con un enlace de registro, que se debe utilizar si el objetivo 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, que se debe utilizar si el objetivo es una respuesta de categoría múltiple. Se utiliza un enlace logit acumulado (resultados ordinales) o un enlace logit generalizado (respuestas nominales con categorías múltiples).
- **Regresión logística binaria.** Especifica una distribución binomial con un enlace logit, que debe utilizarse cuando el objetivo es una respuesta binaria predicha por un modelo de regresión logística.
- **Probit binario.** Especifica una distribución binomial con un enlace probit, que debe utilizarse cuando el objetivo es una respuesta binaria con una distribución normal subyacente.
- **Supervivencia censurada por intervalos.** Especifica una distribución binomial con un enlace log-log complementario, que resulta de utilidad en el análisis de supervivencia cuando algunas observaciones no tienen suceso de terminación.

Distribución

Esta selección especifica la distribución del objetivo. 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 mixto generalizado respecto al modelo lineal mixto. 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 objetivo que represente una respuesta binaria o un número de sucesos.
- **Gamma.** Esta distribución es adecuada para un objetivo 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 objetivo 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 objetivo que representa una respuesta de categorías múltiples. La forma del modelo dependerá del nivel de medición del objetivo.

Un objetivo **nominal** dará como resultado un modelo multinomial nominal en el que se calcula un conjunto independiente de parámetros del modelo para cada categoría del objetivo (excepto la categoría de referencia). Las estimaciones de parámetros de un predictor determinado muestran la relación entre ese predictor y la verosimilitud de cada categoría del objetivo, relativa a la categoría de referencia.

Un objetivo **ordinal** dará como resultado un modelo multinomial ordinal en el que el término de interceptación tradicional se sustituye por un conjunto de parámetros de **umbral** que se relacionan con la probabilidad acumulada de las categorías objetivo.

- **Binomial negativa.** La regresión binomial negativa utiliza una distribución binomial negativa con un enlace de registro, que se debe utilizar cuando el objetivo representa un recuento de apariciones con una varianza alta.
- **Normal.** Es adecuada para un objetivo continuo cuyos valores adoptan una distribución simétrica con forma de campana en torno a un valor central (la media).
- **Poisson.** Esta distribución se puede considerar como el número de ocurrencias de un suceso 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 objetivo que permite una estimación del modelo. Están disponibles las opciones siguientes:

- **Identidad.** $f(x)=x$. El destino no se transforma. Este enlace se puede utilizar con cualquier distribución, excepto la multinomial.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Resulta adecuado solamente con la distribución binomial o multinomial.
- **Cauchit.** $f(x) = \tan(\pi(x - 0.5))$. Resulta adecuado solamente con la distribución binomial o multinomial.
- **Registro.** $f(x)=\log(x)$. Este enlace se puede utilizar con cualquier distribución, excepto la multinomial.
- **Complemento de registro.** $f(x)=\log(1-x)$. Resulta apropiada únicamente para la distribución binomial.
- **Logit.** $f(x)=\log(x / (1-x))$. Resulta adecuado solamente con la distribución binomial o multinomial.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Resulta adecuado solamente con la distribución binomial o multinomial.
- **Probit.** $f(x)=\Phi^{-1}(x)$, donde Φ^{-1} es la función de distribución acumulada normal estándar inversa. Resulta adecuado solamente con la distribución binomial o multinomial.
- **Potencia.** $f(x)=x^\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si es $\alpha=0$. α es la especificación del número requerido y debe ser un número real. Este enlace se puede utilizar con cualquier distribución, excepto la multinomial.

Efectos fijos

Los factores de efectos fijos suelen considerarse campos cuyos valores de interés están todos representados en el conjunto de datos y pueden utilizarse para la puntuación. De forma predeterminada, los campos con el rol de entrada predefinido que no se especifican en ninguna otra parte del diálogo se especifican 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.

Indique los efectos en el modelo seleccionando uno o más campos en la lista de orígenes y arrastrándolos a la lista de efectos. El tipo de efecto que se crea depende de la zona activa en la que se suelte la selección.

- **Principal.** Los campos que se suelten aparecen como efectos principales independientes en la parte inferior de la lista de efectos.

- **2 factores.** Todos los pares posibles de los campos que se suelten aparecen como interacciones de 2 factores en la parte inferior de la lista de efectos.
- **3 vías.** Todos los triples posibles de los campos aparecen como interacciones de 3 vías en la parte inferior de la lista de efectos.
- *****. La combinación de todos los campos descartados aparece como una única interacción en la parte inferior de la lista de efectos.

Los botones a la derecha del creador de efectos le permiten realizar varias acciones.

Tabla 1. Descripciones de los botones del creador de efectos

Icono	Descripción
	Se suprimen los términos del modelo de efectos fijos seleccionando los términos que se desean eliminar y pulsando en el botón Suprimir.
	Se reordenan los términos dentro del modelo de efectos fijos seleccionando los términos que se quieren reordenar y pulsando en la flecha arriba o abajo.
	
	Añadir términos anidados al modelo utilizando el diálogo “Añadir un término personalizado” en la página 57, pulsando el botón Añadir un término personalizado.

Incluir intersección. La intersección se incluye normalmente en el modelo. Si se puede dar por supuesto la lectura de datos en el origen, se puede excluir la interceptación.

Añadir un término personalizado

En este procedimiento puede generar términos anidados para su modelo. Los términos anidados son útiles para modelar el efecto de un factor o covariable cuyos valores no interactúen 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 solamente 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. Los términos anidados tienen las siguientes restricciones:

- Todos los factores incluidos en una interacción deben ser exclusivos. Por consiguiente, si A es un factor, no es válido especificar $A*A$.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos. 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 $A(X)$.

Creación de un término anidado

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

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

Efectos aleatorios

Los factores de efectos aleatorios son campos cuyos valores en el archivo de datos se pueden considerar una muestra aleatoria a partir de una gran población de valores. Resultan útiles para explicar el exceso de variabilidad en el destino. De forma predeterminada, si ha seleccionado más de un asunto en la pestaña Estructura de datos, se creará un bloque de efectos aleatorios para cada asunto más allá del asunto más interior. Por ejemplo, si ha seleccionado Colegio, Clase y Alumno como asuntos en la pestaña Estructura de datos, se crearán los siguientes bloques de efectos aleatorios:

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

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

1. Para añadir un bloque nuevo, pulse **Añadir bloque...** De este modo se abre el diálogo “Bloque de efectos aleatorios” en la página 58.
2. Para editar un bloque existente, seleccione el bloque que desea editar y pulse **Editar bloque ...** De este modo se abre el diálogo “Bloque de efectos aleatorios” en la página 58.
3. Para eliminar uno o más bloques, seleccione los bloques que desee eliminar y pulse el botón Eliminar.

Bloque de efectos aleatorios

Introduzca los efectos en el modelo seleccionando uno o más campos en la lista de origen y añadiéndolos a la lista **Generador de efectos**.

El tipo de efecto creado depende del **Tipo** que seleccione. Los campos categóricos (, nominal y ordinal) se utilizan como factores en el modelo y los campos continuos se utilizan como covariables.

Interacción

La combinación de todos los campos aparece como una única interacción en la parte inferior de la lista de efectos.

Efectos principales

Los campos que se suelten aparecen como efectos principales independientes en la parte inferior de la lista de efectos.

Todos los de 2 factores

Todos los pares posibles de los campos que se suelten aparecen como interacciones de 2 factores en la parte inferior de la lista de efectos.

Todos los de 3 factores

Todos los triples posibles de los campos aparecen como interacciones de 3 factores en la parte inferior de la lista de efectos.

Todos los de 4 factores

Todos los triples posibles de los campos descartados aparecen como interacciones de 4 factores en la parte inferior de la lista de efectos.

Todos los de 5 factores

Todos los triples posibles de los campos descartados aparecen como interacciones de 5 factores en la parte inferior de la lista de efectos.

Incluir intersección

La intersección no está incluida en el modelo de efectos aleatorios de forma predeterminada. Si se puede dar por supuesto la lectura de datos en el origen, se puede excluir la interceptación.

Mostrar predicciones de parámetro para este bloque

Especifica que se visualicen las estimaciones de los parámetros de efectos aleatorios.

Combinación de asuntos

Esto le permite especificar asuntos de efecto aleatorio a partir de combinaciones preestablecidas de asuntos del diálogo Variables. Por ejemplo, si *School*, *Class* y *Student* están definidos como sujetos en el diálogo Variables, y en ese orden, la lista desplegable Combinación de sujetos tendrá **None**, **School**, **School * Class** y **School * Class * Student** como opciones.

Tipo de covarianza de efectos aleatorios

Esto especifica la estructura de covarianzas 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

Definir grupos de covarianzas por

Los campos categóricos especificados aquí definen conjuntos independientes de parámetros de covarianza de efectos aleatorios; uno para cada categoría definida por la clasificación cruzada de los campos de agrupación. Se puede especificar un conjunto diferente de campos de agrupación para cada bloque de efectos aleatorios. Todos los sujetos tienen el mismo tipo de covarianza; los sujetos con la misma agrupación de covarianza tendrán los mismos valores para los parámetros.

Ponderación y desplazamiento

Ponderación de análisis. El parámetro de escala es un parámetro del modelo estimado relacionado con la varianza de la respuesta. Las ponderaciones 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 entre los valores de ponderación de análisis para cada observación. Los registros con valores de ponderación de análisis que sean inferiores o iguales a 0 o que sean valores perdidos no se utilizarán en el análisis.

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 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 Poisson o una binomial negativa con un enlace de registro si el registro natural de la experiencia del conductor se incluye como un término de desplazamiento.

Otras combinaciones de los tipos de distribución y enlace requerirán otras transformaciones de la variable de desplazamiento.

Opciones de generación generales

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" en la página 54.

Reglas de parada

Puede especificar el número máximo de iteraciones que ejecutará el algoritmo. El algoritmo utiliza un proceso iterativo doble que consta de un bucle interno y un bucle externo. El valor especificado para el número máximo de iteraciones se aplica a ambos bucles. Especifique un número entero no negativo. El valor predeterminado 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 predeterminado es 95.

Grados de libertad

Especifica cómo se calculan los grados de libertad para las comprobaciones de significación. Elija **Método residual** si el tamaño de su muestra es suficientemente grande, si los datos están equilibrados o si el modelo utiliza un tipo de covarianza más simple (por ejemplo, identidad escalada o diagonal). Ésta es el ajuste predeterminado. Elija **Aproximación de Satterthwaite** si el tamaño de su muestra es pequeño, si los datos no están equilibrados o si el modelo utiliza un tipo de covarianza complicado (por ejemplo, sin estructura). Elija **Aproximación de Kenward-Roger** si el tamaño de la muestra es pequeño y tiene un modelo Máxima verosimilitud restringida (REML).

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.

Estimación

El algoritmo de generación del modelo utiliza un proceso iterativo doble que consta de un bucle interior y un bucle exterior. Los valores siguientes se aplican al bucle interno.

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 del logaritmo de verosimilitud.

Se asume la convergencia si el cambio absoluto o el cambio relativo en la función de log-verosimilitud es menor que el valor especificado, que no debe ser 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.

Pasos máximos de puntuación de Fisher.

Especifique un número entero no negativo. Un valor de 0 especifica el método Newton-Raphson. Los valores mayores que 0 especifican el uso del algoritmo de puntuación de Fisher hasta el número de iteración n , donde n es el entero especificado y después de Newton-Raphson.

Tolerancia para la singularidad.

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

Nota: De forma predeterminada, se utiliza la convergencia de parámetros, donde se marca el cambio **Absoluto** máximo a una tolerancia de 1E-6. Este valor podría generar resultados que difieren de los resultados obtenidos en versiones anteriores a la versión 22. Para reproducir los resultados de versiones previas a las 22, utilice **Relativo** para el criterio Convergencia de parámetro y mantenga el valore de convergencia predeterminado de 1E-6.

Medias estimadas

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

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

- **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 para todas las combinaciones de niveles de los factores especificados. Este contraste es el único disponible para las interacciones de los factores. **Contrastes de desviación** comparan 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 las Opciones de creación. Tenga en cuenta que todos estos tipos de contrastes no son ortogonales.
- **Campo de contraste.** Esto especifica un factor, cuyos niveles se comparan mediante 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 Efectos fijos que utilizan campos continuos. Al calcular medias marginales estimadas, las covariables están fijas en los valores especificados. Seleccione la media o especifique un valor personalizado.

Mostrar medias estimadas según. Esto especifica si las medias marginales estimadas se calculan basándose en la escala original del objetivo o basándose en la transformación de la función de enlace. **Escala original del objetivo** calcula las medias marginales del destino. Tenga en cuenta que cuando el objetivo se especifica mediante la opción de sucesos/pruebas, proporciona la media marginal estimada de la proporción de sucesos/pruebas en lugar de la del número de sucesos. **Transformación de la función de enlace** calcula la media marginal estimada 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. Esto 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.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Sidak secuencial.* Éste 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 Sidak secuencial, que a su vez es menos conservador que Bonferroni secuencial; es decir, la diferencia menos significativa rechazará al menos tantas hipótesis individuales como Sidak secuencial, que a su vez rechazará al menos tantas hipótesis individuales como Bonferroni secuencial.

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 predicho del destino. El nombre del campo predeterminado 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**. Los valores calculados son probabilidades acumuladas para destinos ordinales. El nombre de raíz predeterminado 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 predicho o la probabilidad pronosticada. Para todas las distribuciones excepto la multinomial, crea dos variables y el nombre de raíz predeterminado es *CI*, con *_Lower* y *_Upper* como sufijos.

Para la distribución multinomial y un destino nominal, se crea un campo para cada categoría de variable dependiente. Esta guarda los límites inferior y superior de la probabilidad pronosticada de las primeras *n* categorías, hasta el valor especificado como **Máximo de categorías para guardar**. El nombre de raíz predeterminado es *CI* y los nombres de campos predeterminados 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.

Para la distribución multinomial y un destino ordinal, se crea un campo para cada categoría de variable dependiente excepto la última (consulte el tema "[Opciones de generación generales](#)" en la [página 59](#) para obtener más información). 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 predeterminado es *CI* y los nombres de campos predeterminados 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 predeterminado es *ResiduoPearson*.

Confianzas. Guarda la confianza en el valor predicho del destino categórico. La confianza calculada puede basarse en la probabilidad del valor predicho (la probabilidad predicha más alta) o la diferencia entre la probabilidad predicha más alta y la segunda probabilidad predicha más alta. El nombre del campo predeterminado 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. para obtener más información. Especifique un nombre de archivo exclusivo y válido. Si la especificación del archivo hace referencia a un archivo existente, se sobrescribirá el archivo.

Vista de modelo

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

De forma predeterminada, se muestra la vista Resumen del modelo. Para ver otra vista de modelo, selecciónela entre las vistas en miniatura.

Como alternativa al objeto Modelo, puede generar tablas dinámicas y gráficos seleccionando **Tablas dinámicas y gráficos** en el grupo Visualización de salida en la pestaña Salida del diálogo Opciones (Editar > Opciones). Los temas que siguen describen el objeto Modelo.

Resumen del modelo

Esta vista es una instantánea, un resumen detallado 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 objetivo se define mediante sucesos y ensayos, la celda se divide para mostrar el campo de sucesos y el campo de ensayos o el número fijo de ensayos. Además, se muestran el criterio de información de Akaike (AICC) corregido para muestras finitas y el criterio de información bayesiano (BIC).

- *Akaike Corregido.* Una medida para seleccionar y comparar modelos mixtos basada en la log-verosimilitud -2 (restringida). Los valores menores indican modelos mejores. El AICC "corrige" el AIC respecto a tamaños muestrales pequeños. A medida que aumenta el tamaño de la muestra, el AICC converge con el AIC.
- *Bayesiano.* Una medida para seleccionar y comparar modelos basados en el logaritmo de la verosimilitud -2. Los valores menores indican modelos mejores. El BIC también "penaliza" modelos sobrep parametrizados (modelos complejos con un gran número de entradas, por ejemplo), pero de forma 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

Esta vista proporciona un resumen de la estructura de datos que especifique y le ayuda a comprobar que los sujetos y las medidas repetidas se han especificado correctamente. La información observada para el primer sujeto se muestra para cada campo de sujeto y campo de medidas repetidas, así como el objetivo. Además, se muestra el número de niveles de cada campo de sujeto y campo de medidas repetidas.

Predicho por observado

En los destinos continuos, incluyendo los destinos especificados como eventos/ensayos, se muestra un diagrama de dispersión en intervalos de los valores pronosticados en el eje vertical y de los valores observados en el eje horizontal. Idealmente, los puntos deben basarse en una línea de 45 grados; esta vista indica si hay algún registro predicho de manera incorrecta en el modelo.

Clasificación

Para los objetivos categóricos, muestra la clasificación cruzada de los valores observados en contraposición a los predichos en el mapa de calor, junto con 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. Este es el valor predeterminado.
- **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 cuando el objetivo tiene muchas 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.** Éste es un gráfico en el que los efectos están ordenados de arriba a abajo en el orden en que se especificaron 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). Este es el valor predeterminado.
- **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 especificaron 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, sino que simplemente permite centrarse en los efectos más importantes. El valor predeterminado es 1,00, de modo que no se filtran efectos tomando como base la significación.

Coefficientes 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 predeterminado.
- **Tabla.** Muestra los valores, las pruebas de significación y los intervalos de confianza para los coeficientes de modelos individuales. Después de la interceptación, los efectos están ordenados de arriba a abajo en el orden en que se especificaron 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 viene determinado por la especificación de la configuración de Opciones de generación.

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 predeterminado 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.** Éste es un mapa de calor de la matriz de covarianzas en el que los efectos están ordenados de arriba a abajo en el orden en que se especificaron en la configuración de Efectos fijos. Los colores de corrgram se corresponden con los valores de la celda tal y como se muestran en la clave. Este es el valor predeterminado.
- **Corrgram.** Es un mapa de calor de la matriz de covarianzas.
- **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 efectos aleatorios tiene una especificación de grupo, existe una lista desplegable Grupo para seleccionar el nivel de grupo que se muestra.

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 viene determinado por la especificación de la configuración de Opciones de generación.

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 Ésta es una referencia rápida al número de parámetros en las matrices de covarianza de efectos residuales (**R**) y aleatorios (**G**), el rango (número de columnas) en las matrices de diseño de efectos fijos (**X**) y efectos aleatorios (**Z**) y el número de sujetos definidos por los campos de sujeto que definen la estructura de datos.

Tabla de parámetros de covarianza. Para el efecto seleccionado, la estimación, error estándar y el intervalo de confianza se muestran para cada parámetro de covarianza. El número de parámetros que se muestra depende de la estructura de la covarianza del efecto y, en el caso de bloques de efectos aleatorios, el número de efectos del bloque. Si observa que los parámetros de fuera de la diagonal no son significativos, tal vez pueda 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 siempre está 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 viene determinado por la especificación de la configuración de Opciones de generación.

Medias estimadas: Efectos significativos

Se trata de gráficos que se muestran para los 10 efectos de todos los factores fijos "más significativos", comenzando por las interacciones de 3 factores, seguidas de las interacciones de 2 factores y, por último, de los efectos principales. El gráfico muestra el valor estimado por el modelo del objetivo en el eje vertical para cada valor del efecto principal (o el primer efecto enumerado en una interacción) en el eje horizontal; se genera una línea independiente para cada valor del segundo efecto enumerado en una interacción; se genera un gráfico independiente para cada valor del tercer efecto enumerado en una interacción de 3 factores; el resto de predictores se mantiene constante. 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 generación.

Medias estimadas: Efectos personalizados

Estas son tablas y gráficos para efectos de todos los factores fijos solicitados por los usuarios.

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

- **Diagrama.** Este estilo muestra un gráfico de líneas del valor estimado por el modelo del objetivo en el eje vertical para cada valor del efecto principal (o el primer efecto enumerado en una interacción) en el eje horizontal; se genera una línea independiente para cada valor del segundo efecto enumerado en una interacción; se genera un gráfico independiente para cada valor del tercer efecto enumerado en una interacción de 3 factores; el resto de predictores se mantiene constante.

Si se solicitan contrastes, se muestra otro gráfico para comparar los niveles del campo Contraste; para las interacciones, se muestra un gráfico para cada combinación de niveles de los efectos distintos del campo 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 se corresponden con diferencias significativas estadísticamente; las líneas negras se corresponden con diferencias no significativas. Al pasar el ratón por encima de una

línea de la red, aparece información sobre herramientas con el significado ajustado de la diferencia entre los nodos conectados por la línea.

Para **Contrastes de desviación**, se muestra un gráfico de barras con el valor estimado por el modelo del objetivo en el eje vertical y los valores del campo Contraste en el eje horizontal; para las interacciones, se muestra un gráfico para cada combinación de niveles de los efectos distintos del campo Contraste. Las barras muestran la diferencia entre cada nivel del campo Contraste y la media global, que está representada 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 Contraste (excepto el último) y el último nivel, que está representado por una línea horizontal negra.

- **Tabla.** Este estilo muestra una tabla de valores del destino estimados por el usuario, su error estándar 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 solicitan contrastes, se muestra otra tabla con la estimación, el error estándar, la prueba de significación y el intervalo de confianza para cada contraste; para las interacciones, hay un conjunto de filas independiente para cada combinación de niveles de los efectos distintos del campo Contraste. Además, se muestra una tabla con los resultados de las pruebas globales; para las interacciones, hay una prueba global independiente para cada combinación de niveles de los efectos distintos del campo 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 generación.

Diseño. Esto cambia el diseño del diagrama de contrastes por parejas. El diseño circular muestra menos de los contrastes que el diseño de red, pero evita que se superpongan las líneas.

Análisis loglineal: Selección de modelo

El procedimiento de análisis loglineal de selección de modelo analiza tablas cruzadas (tablas de contingencia) de varios factores. Ajusta modelos loglineales jerárquicos a las tabulaciones cruzadas multidimensionales utilizando un algoritmo de ajuste proporcional. Este procedimiento ayuda a encontrar cuáles de las variables categóricas están asociadas. Para generar 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 estándar, intervalos de confianza y pruebas de asociación parcial. Para los modelos personalizados, gráficos de residuos y gráficos de probabilidad normal.

Consideraciones sobre los datos de Análisis loglineal: Selección de modelo

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...

1. Seleccione dos o varios factores categóricos numéricos.
2. Seleccione una o más variables de factor en la lista Factores y pulse en **Definir rango**.
3. Defina el rango de valores para cada variable de factor.
4. Seleccione una opción en la sección Generación de modelos.

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

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

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 Crear términos. El modelo resultante contendrá la interacción de 3 vías especificada $A*B*C$; las interacciones de 2 vías $A*B$, $A*C$ y $B*C$, y los principales efectos para A , B y C . No especifique los relativos de orden inferior en la clase generadora.

Para los factores seleccionados:

Interacción

Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales

Crea un término de efectos principales para cada variable seleccionada.

Todos los de 2 factores

Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todos los de 3 factores

Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todos los de 4 factores

Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todos los de 5 factores

Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

Construcción de términos y Términos personalizados

Construir términos

Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados

Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

Análisis loglineal: Opciones

Visualización. 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 **Nº máximo de iteraciones**, **Convergencia** o **Delta** (un valor añadido a todas las frecuencias de casilla para los modelos saturados).

Características adicionales del comando HILOGLINEAR

La sintaxis de comandos también le permite:

- 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).

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

Análisis loglineal general

El procedimiento Análisis loglineal general analiza las frecuencias de las observaciones incluidas en cada categoría de la clasificación cruzada de una tabulación cruzada o 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 tabulación cruzada 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 de las log-odds ratio 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.

Análisis loglineal general: Consideraciones sobre los datos

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 las log-razones de las ventajas generalizadas. Los valores de la variable de contraste son los coeficientes para la combinación lineal de los logaritmos de los recuentos de casillas esperados.

Una variable de estructura de casilla asigna ponderaciones. Por ejemplo, si algunas de las casillas son ceros estructurales, la variable de estructura de casilla tiene un valor de 0 o 1. No utilice una variable de estructura de casilla para ponderar 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 otras casillas.

Bajo el supuesto de distribución multinomial:

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

Procedimientos relacionados. Utilice el procedimiento Tablas cruzadas para examinar las tabulaciones cruzadas. 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

1. En los menús seleccione:

Analizar > Loglineal > General...

2. 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

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.

Para las covariables y los factores seleccionados:

Interacción

Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales

Crea un término de efectos principales para cada variable seleccionada.

Todos los de 2 factores

Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todos los de 3 factores

Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todos los de 4 factores

Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todos los de 5 factores

Crea todas las interacciones quíntuples posibles de las variables seleccionadas.

Construcción de términos y Términos personalizados

Construir términos

Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados

Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

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:

Representación. 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. También puede 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. El método Newton-Raphson se utiliza para obtener las estimaciones de parámetro de máxima verosimilitud. Se pueden especificar 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

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 casillas.

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

- *Residuos.* También llamado residuo simple o bruto, es la diferencia entre el recuento de casilla observado y el esperado.

- *Residuos estandarizados*. Los residuos divididos por una estimación de su error estándar. Los residuos estandarizados se conocen también como residuos de Pearson.
- *Residuos ajustados*. El residuo estandarizado dividido por la estimación de su error estándar. Dado que, cuando el modelo es el correcto, los residuos corregidos son asintóticamente normales estándar, éstos son preferidos a los residuos estandarizados a la hora de contrastar la normalidad.
- *Residuos de desviación*. 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 (recuento observado menos recuento esperado). Los residuos de desviación tienen una distribución normal estándar asintótica.

Características adicionales del comando GENLOG

La sintaxis de comandos también le permite:

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

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

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 la probabilidad 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 de las log-odds ratio generalizadas (GLOR). Los valores de la variable de contraste son los coeficientes para la combinación lineal de los logaritmos de los recuentos de casillas esperados.

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 probabilidad de que un caimán pequeño prefiera reptiles a peces es 0,70 veces menor que la de un caimán grande; además la probabilidad 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 ventajas generalizadas; estadístico de Wald; intervalos de confianza. Gráficos: residuos corregidos, residuos de desviación y gráficos de probabilidad normal.

Análisis loglineal logit: Consideraciones sobre los datos

Datos. Las variables dependientes son categóricas. Los factores son categóricos. 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 esa casilla. Las variables de contraste son continuas. Se utilizan para calcular los logaritmos de la razón de probabilidad generalizada (GLOR). Los valores de la variable de contraste son los coeficientes para la combinación lineal de los logaritmos de los recuentos de casillas esperados.

Una variable de estructura de casilla asigna ponderaciones. Por ejemplo, si algunas de las casillas son ceros estructurales, la variable de estructura de casilla tiene un valor de 0 o 1. No utilice una variable de estructura de casilla para ponderar datos agregados. 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 total de la muestra es fijo o el análisis está condicionado al tamaño de la muestra total.
- Los recuentos de casillas no son estadísticamente independientes.

Procedimientos relacionados. Utilice el procedimiento Tablas cruzadas para mostrar las tablas de contingencia. Utilice el procedimiento Análisis loglineal general cuando quiera analizar la relación entre un recuento observado y un conjunto de variables explicativas.

Para obtener un análisis loglineal logit

1. En los menús seleccione:

Analizar > Loglineal > Logit...

2. En el cuadro de diálogo Análisis loglineal logit, seleccione una o más variables dependientes.
3. 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

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.

Para las covariables y los factores seleccionados:

Interacción

Crea el término de interacción de mayor nivel con todas las variables seleccionadas. Este es el valor predeterminado.

Efectos principales

Crea un término de efectos principales para cada variable seleccionada.

Todos los de 2 factores

Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todos los de 3 factores

Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todos los de 4 factores

Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todos los de 5 factores

Crea todas las interacciones quintuples posibles de las variables seleccionadas.

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.

Construcción de términos y Términos personalizados

Construir términos

Utilice esta opción si desea incluir términos no anidados de un tipo determinado (por ejemplo, efectos principales) para todas las combinaciones de un conjunto seleccionado de factores y covariables.

Construir términos personalizados

Utilice esta opción si desea incluir términos anidados o si desea crear explícitamente una variable de término por variable. La creación de un término anidado implica los pasos siguientes:

Análisis loglineal logit: Opciones

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

Visualización. 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 disponibles para los modelos personalizados incluyen dos diagramas de dispersión matriciales (los residuos corregidos o los residuos de desviación respecto a los recuentos de casillas observados y esperados). También puede 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 máximo-verosímiles de los parámetros. Se pueden especificar 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

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 casillas.

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 el recuento de casilla observado y el esperado.
- *Residuos estandarizados*. Los residuos divididos por una estimación de su error estándar. Los residuos estandarizados se conocen también como residuos de Pearson.
- *Residuos ajustados*. El residuo estandarizado dividido por la estimación de su error estándar. Dado que, cuando el modelo es el correcto, los residuos corregidos son asintóticamente normales estándar, éstos son preferidos a los residuos estandarizados a la hora de contrastar la normalidad.
- *Residuos de desviación*. 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 (recuento observado menos recuento esperado). Los residuos de desviación tienen una distribución normal estándar asintótica.

Características adicionales del comando GENLOG

La sintaxis de comandos también le permite:

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

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

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. Esto puede producirse por distintas razones: 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 riesgo 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 estándar), densidad de probabilidad (y error estándar), tasa de riesgo (y error estándar) para cada intervalo de tiempo en cada grupo. Gráficos: gráficos de las funciones para supervivencia, log de la supervivencia, densidad, tasa de riesgo y uno menos la supervivencia.

Tablas de mortalidad: Consideraciones sobre los datos

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

1. En los menús seleccione:

Analizar > Superviv. > Tablas de mortalidad...

2. Seleccione una variable *numérica* de supervivencia.

3. Especifique los intervalos de tiempo que se van a examinar.

4. Seleccione una variable de estado para definir casos para los que tuvo lugar el evento terminal.

5. 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

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. Especifique un solo valor o un rango de valores que identifiquen el evento de interés.

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 exclusivo dentro del rango.

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, riesgo, 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.
- *Riesgo.* Muestra la función de riesgo acumulado 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.

Características adicionales del comando SURVIVAL

La sintaxis de comandos también le permite:

- 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.

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

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 riesgo o de supervivencia y compararlas visualmente.

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

Kaplan-Meier: Consideraciones sobre los datos

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 riesgo 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.

Para obtener un análisis de supervivencia de Kaplan-Meier

1. En los menús seleccione:

Analizar > Superviv. > Kaplan-Meier...

2. Seleccione una variable de tiempo.

3. 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

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

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.

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.

- *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 pareja de niveles del factor diferente. 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.

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 estándar de la supervivencia, el riesgo y los eventos acumulados, como nuevas variables.

- *Supervivencia*. Estimación de la probabilidad de supervivencia acumulada. El nombre de variable predeterminado 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 estándar de supervivencia*. Error estándar de la estimación de la supervivencia acumulada. El nombre de variable predeterminado es el prefijo se_ con un número secuencial. Por ejemplo, si se_1 ya existe, Kaplan-Meier asigna el nombre de variable se_2.
- *Riesgo*. Estimación de la función de riesgo acumulada. El nombre de variable predeterminado 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 predeterminado 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

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, riesgo 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.
- *Riesgo*. Muestra la función de riesgo acumulado en una escala lineal.
- *Log de la supervivencia*. Muestra la función de supervivencia, en una escala logarítmica.

Características adicionales del comando KM

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.
- Especificar el espaciado desigual para la prueba de la tendencia lineal.
- Obtener percentiles diferentes a los cuartiles para la variable del tiempo de supervivencia.

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

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 predice 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 mediciones para las variables predictoras. Observe que la información de los sujetos 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 estándar y Estadísticos de Wald. Para variables que no estén en el modelo: Estadísticos de puntuación y Chi-cuadrado residual.

Regresión de Cox: Consideraciones sobre los datos

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 riesgo debe ser constante a lo largo del tiempo; es decir, la proporcionalidad de los riesgos de un caso a otro no debe variar en función del tiempo. El último supuesto se conoce como el **supuesto de proporcionalidad de los riesgos**.

Procedimientos relacionados. Si el supuesto de proporcionalidad de los riesgos no se cumple (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 riesgo 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

1. En los menús seleccione:

Analizar > Supervivencia > Regresión de Cox...

2. Seleccione una variable de tiempo. No se analizan aquellos casos en los que los valores del tiempo son negativos.
3. Seleccione una variable de estado y pulse en **Definir evento**.
4. Seleccione una o varias covariables. Para incluir términos de interacción, seleccione todas las variables implicadas en la interacción y, a continuación, pulse **>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

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 muévela 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 contrastes como una fila de ceros.
- **Simples.** 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.** Contrastos 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

Los gráficos pueden ayudarle a evaluar el modelo estimado e interpretar los resultados. Es posible representar gráficamente las funciones de supervivencia, de riesgo, log-menos-log y uno menos la supervivencia.

Tipo de gráfico

Superviv.

Muestra la función de supervivencia acumulada, en una escala lineal.

Riesgo

Muestra la función de riesgo acumulado en una escala lineal.

Log menos log

La estimación de la supervivencia acumulada tras la transformación $\ln(-\ln)$ se aplica a la estimación.

Uno menos la supervivencia

Representa la función uno menos la supervivencia en una escala lineal.

Valores de covariable representados en

Puede trazar una línea separada para cada valor de una covariable categórica moviendo dicha covariable el recuadro de texto **Líneas separadas para**. Esta opción solo está disponible para covariables categóricas, que se denotan mediante **(Cat)** detrás de sus nombres en la lista **Valores de covariable trazados en**.

Cambiar valor

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 predeterminado es utilizar la media de cada covariable como un valor constante, pero puede especificar sus propios valores para el gráfico utilizando el grupo de control **Valor de cambio**.

Regresión de Cox: Guardar variables nuevas

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 estándar, estimaciones de log menos log, función de riesgo, residuos parciales, DfBeta(s) de la regresión y predictor lineal $X \cdot \text{Beta}$ 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.

Error estándar de la función de supervivencia

Error estándar de la estimación de supervivencia acumulada

Log menos log de la función de supervivencia

La estimación de la supervivencia acumulada tras la transformación $\ln(-\ln)$ se aplica a la estimación.

Función de riesgo

Guarda la estimación de función de riesgo acumulada (también llamado el residuo de Cox-Snell).

Residuos parciales

Puede representar los residuos parciales respecto al tiempo para probar el supuesto de proporcionalidad de los riesgos. 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.

$X \cdot \text{Beta}$

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.

Nota: Solo se guardan las DfBeta(s) si está ejecutando Cox con una covariable dependiente del tiempo.

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

Es posible controlar diferentes aspectos del análisis y los resultados.

Estadísticos del modelo. Puede obtener estadísticos para sus parámetros del modelo, incluidos los intervalos de confianza para $\exp(B)$ y correlaciones de estimaciones. Puede solicitar estos estadísticos en cada paso o sólo en el último paso.

Probabilidad para el método por pasos. Si ha seleccionado un método escalonado, puede especificar la probabilidad de entrada o de salida 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 Eliminación. El valor de Entrada debe ser menor que el valor de Eliminación.

Nº máximo de iteraciones. Permite especificar las iteraciones máximas del modelo, que controla durante cuánto tiempo el procedimiento buscará una solución.

Mostrar la función de línea base. Permite mostrar la función de riesgo de la línea de base 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.

Características 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 archivos segmentados en un archivo de trabajo 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.

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 proporcionalidad de los riesgos. Es decir, que las tasas de riesgo 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 denomina $T_.$ Puede utilizar esta variable para definir covariables dependientes del tiempo de dos formas generales:

- Si desea probar el supuesto de proporcionalidad de los riesgos con respecto a una covariable particular o bien desea estimar un modelo de regresión de Cox extendido que permita riesgos 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 riesgos.

- 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 mide la presión arterial una vez a la semana durante las cuatro semanas de su estudio (identificado como *BP1* a *BP4*), puede definir su covariable dependiente del tiempo como $(T_ < 1) * BP1 + (T_ >= 1 \& T_ < 2) * BP2 + (T_ >= 2 \& T_ < 3) * BP3 + (T_ >= 3 \& T_ < 4) * BP4$. Observe que exactamente uno de los términos entre paréntesis será igual a 1 para cualquier caso dado y el resto será igual a 0. En otras palabras, esta función significa que si el tiempo es inferior a una semana, se utiliza *BP1*; si es más de una semana pero menos de dos semanas, se utiliza *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

1. En los menús seleccione:

Analizar > Supervivencia > Cox w/ Time-Dep Cov...

2. Introduzca una expresión para la covariable dependiente del tiempo.

3. 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.

Consulte el tema “Análisis de regresión de Cox” en la página 79 para obtener más información.

Regresión de Cox con covariables dependientes del tiempo: Características adicionales

El lenguaje de sintaxis de comandos permite especificar múltiples covariables dependientes del tiempo. Dispone además de otras características de sintaxis de comandos para la regresión de Cox con o sin covariables dependientes del tiempo.

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

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:

```
mean ( 1/k 1/k ... 1/k 1/k)
df(1) (1-1/k -1/k ... -1/k -1/k)
df(2) (-1/k 1-1/k ... -1/k -1/k)
```

```
df(k-1) ( -1/k  -1/k  ... 1-1/k  -1/k)
```

donde k es el número de categorías para la variable independiente y, de forma predeterminada, 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 contrastes resultante será

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3  2/3)
```

Sencillo

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

```
mean (1/k  1/k  ...  1/k  1/k)
df(1) ( 1  0  ...  0  -1)
df(2) ( 0  1  ...  0  -1)
.
.
df(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 contrastes que omite la segunda categoría:

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

Suponga que *factor* tiene cuatro categorías. La matriz de contrastes 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

```
mean (1/k  1/k  ...  1/k  1/k  1/k)
df(1) ( 1  -1/(k-1)  ...  -1/(k-1)  -1/(k-1)  -1/(k-1))
df(2) ( 0  1  ...  -1/(k-2)  -1/(k-2)  -1/(k-2))
.
.
df(k-2) ( 0  0  ...  1  -1/2  -1/2)
df(k-1) ( 0  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 contrastes 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

```
mean ( 1/k  1/k  1/k  ...  1/k)
df(1) ( -1   1   0  ...  0)
df(2) ( -1/2 -1/2  1  ...  0)
      :
df(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 predeterminado 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

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      .
      .
df(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 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) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0 2 -1 -1) compare 2nd with 3rd and 4th
(0 0 1 -1) compare 3rd with 4th
```

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 medias suma 0. Los productos de cada par de filas separadas también suman 0:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 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 en la categoría de referencia se codifican como 0 para todas las variables $k-1$. 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.

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.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_4\sigma_1\rho_1\rho_2\rho_3)$
$(\sigma_2\sigma_1\rho_1$	σ_2^2	$\sigma_3\sigma_2\rho_2$	$\sigma_4\sigma_2\rho_2\rho_3)$
$(\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_3\sigma_2\rho_2$	σ_3^2	$\sigma_4\sigma_3\rho_3)$
$(\sigma_4\sigma_1\rho_1\rho_2\rho_3$	$\sigma_4\sigma_2\rho_2\rho_3$	$\sigma_4\sigma_3\rho_3$	$\sigma_4^2)$

AR(1). Se trata de una estructura autorregresiva de primer orden con varianzas homogéneas. La correlación entre dos elementos cualesquiera es igual a ρ para elementos adyacentes, ρ^2 para elementos separados por un tercero, y así sucesivamente y está restringido de modo que $-1 < \rho < 1$.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3)$
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho^2)$
$(\sigma^2\rho^2$	$\sigma^2\rho$	σ^2	$\sigma^2\rho)$
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2)$

AR(1): Heterogénea. Se trata de una estructura autorregresiva de primer orden con varianzas heterogéneas. La correlación entre dos elementos cualesquiera es igual a r para los elementos adyacentes, r^2 para dos elementos separados por un tercero, y así sucesivamente y está limitado a estar entre -1 y 1 .

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3)$
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2)$
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho)$
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	$\sigma_4^2)$

ARMA(1,1). Se trata de una estructura de media móvil autorregresiva. Tiene varianzas homogéneas. La correlación entre dos elementos es igual a ϕ para los elementos adyacentes, ϕ^2 para los elementos separados por un tercero, y así sucesivamente y son los parámetros de promedio autorregresivo y móvil, respectivamente, y sus valores están limitados a estar entre -1 y 1 , ambos inclusive.

$(\sigma^2$	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho^3)$
$(\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2)$
$(\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho)$

$$\begin{pmatrix} \sigma^2\varphi\rho^3 & & & \\ & \sigma^2\varphi\rho^2 & & \\ & & \sigma^2\varphi\rho & \\ & & & \sigma^2 \end{pmatrix}$$

Simetría compuesta. Esta estructura tiene una varianza y una covarianza constantes.

$$\begin{pmatrix} \sigma^2 + \sigma_1^2 & & & \\ \sigma_1 & \sigma^2 + \sigma_1^2 & & \\ \sigma_1 & & \sigma^2 + \sigma_1^2 & \\ \sigma_1 & & & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

Simetría compuesta: Métrica de correlación. Esta estructura de covarianza tiene varianzas y correlaciones homogéneas entre los elementos.

$$\begin{pmatrix} \sigma^2 & & & \\ \sigma^2\rho & \sigma^2 & & \\ \sigma^2\rho & & \sigma^2 & \\ \sigma^2\rho & & & \sigma^2 \end{pmatrix}$$

Simetría compuesta: Heterogénea. Esta estructura de covarianza tiene varianzas heterogéneas y una correlación constante entre los elementos.

$$\begin{pmatrix} \sigma_1^2 & & & \\ \sigma_2\sigma_1\rho & \sigma_2^2 & & \\ \sigma_3\sigma_1\rho & & \sigma_3^2 & \\ \sigma_4\sigma_1\rho & & & \sigma_4^2 \end{pmatrix}$$

Diagonal. Esta estructura de covarianza tiene varianzas heterogéneas y una correlación cero entre los elementos.

$$\begin{pmatrix} \sigma_1^2 & & & \\ 0 & \sigma_2^2 & & \\ 0 & & \sigma_3^2 & \\ 0 & & & \sigma_4^2 \end{pmatrix}$$

AR1 de producto directo (UN_AR1). Especifica el producto Kronecker de una matriz sin estructura y la otra matriz de covarianza de autorregresión de primer orden. La primera matriz sin estructura modela la observación multivariante y la segunda estructura de covarianzas de autorregresión de primer orden modela la covarianza de datos a lo largo del tiempo u otro factor.

Producto directo sin estructura (UN_UN). Especifica el producto Kronecker de dos matrices sin estructura, donde la primera modela la observación multivariante y la segunda modela la covarianza de datos a lo largo del tiempo u otro factor.

Simetría compuesta de producto directo (UN_CS). Especifica el producto Kronecker de una matriz sin estructura y la otra matriz de covarianza de simetría compuesta con varianza constante y covarianza. La primera matriz sin estructura modela la observación multivariante y la segunda estructura de covarianzas de simetría compuesta modela la covarianza de datos a lo largo del tiempo u otro factor.

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 cualesquiera es la raíz cuadrada del producto de sus términos de varianza heterogéneos.

$$\begin{pmatrix} \lambda_1^2 + d & & & \\ & \lambda_2\lambda_1 & & \\ & & \lambda_3\lambda_1 & \\ & & & \lambda_4\lambda_1 \end{pmatrix}$$

$(\lambda_2\lambda_1$	$\lambda_2^2 + d$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2)$
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d$	$\lambda_4\lambda_3)$
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d)$

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 cualesquiera es la raíz cuadrada del producto del primer término de los términos de varianza heterogéneos.

$(\lambda_1^2 + d_1$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1)$
$(\lambda_2\lambda_1$	$\lambda_2^2 + d_2$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2)$
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d_3$	$\lambda_4\lambda_3)$
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d_4)$

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.

$(\sigma_1^2$	$[\sigma_1^2 + \sigma_2^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_2^2]/2 - \lambda$	σ_2^2	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	σ_3^2	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda)$
$([\sigma_1^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$	$\sigma_4^2)$

Identidad escalada. Esta estructura tiene una varianza constante. Se asume que no existe correlación alguna entre los elementos.

$(\sigma^2$	0	0	0)
(0	σ^2	0	0)
(0	0	σ^2	0)
(0	0	0	$\sigma^2)$

Espacial: Potencia. Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre elementos. d_{ij} es la distancia euclídea estimada entre la medición de i^{th} y j^{th} .

$(\sigma^2$	$\sigma^2 \rho^{d_{12}}$	$\sigma^2 \rho^{d_{13}}$	$\sigma^2 \rho^{d_{14}})$
$(\sigma^2 \rho^{d_{12}}$	σ^2	$\sigma^2 \rho^{d_{23}}$	$\sigma^2 \rho^{d_{24}})$
$(\sigma^2 \rho^{d_{13}}$	$\sigma^2 \rho^{d_{23}}$	σ^2	$\sigma^2 \rho^{d_{34}})$
$(\sigma^2 \rho^{d_{14}}$	$\sigma^2 \rho^{d_{24}}$	$\sigma^2 \rho^{d_{34}}$	$\sigma^2)$

Espacial: Exponencial. Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre elementos. d_{ij} es la distancia euclídea estimada entre la medición de i^{th} y j^{th} .

$(\sigma^2$	$\sigma^2 \exp\{-d_{12}/\theta\}$	$\sigma^2 \exp\{-d_{13}/\theta\}$	$\sigma^2 \exp\{-d_{14}/\theta\})$
$(\sigma^2 \exp\{-d_{12}/\theta\}$	σ^2	$\sigma^2 \exp\{-d_{23}/\theta\}$	$\sigma^2 \exp\{-d_{24}/\theta\})$
$(\sigma^2 \exp\{-d_{13}/\theta\}$	$\sigma^2 \exp\{-d_{23}/\theta\}$	σ^2	$\sigma^2 \exp\{-d_{34}/\theta\})$
$(\sigma^2 \exp\{-d_{14}/\theta\}$	$\sigma^2 \exp\{-d_{24}/\theta\}$	$\sigma^2 \exp\{-d_{34}/\theta\}$	$\sigma^2)$

Espacial: De Gauss. Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre elementos. d_{ij} es la distancia euclídea estimada entre la medición de i^{th} y j^{th} .

$(\sigma^2$	$\sigma^2 \exp\{-d_{12}/\rho^2\}$	$\sigma^2 \exp\{-d_{13}/\rho^2\}$	$\sigma^2 \exp\{-d_{14}/\rho^2\})$
-------------	-----------------------------------	-----------------------------------	------------------------------------

$$\begin{array}{cccc}
(\sigma^2 \exp\{-d_{12}/\rho^2\}) & \sigma^2 & \sigma^2 \exp\{-d_{23}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} \\
(\sigma^2 \exp\{-d_{13}/\rho^2\}) & \sigma^2 \exp\{-d_{23}/\rho^2\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\rho^2\} \\
(\sigma^2 \exp\{-d_{14}/\rho^2\}) & \sigma^2 \exp\{-d_{24}/\rho^2\} & \sigma^2 \exp\{-d_{34}/\rho^2\} & \sigma^2
\end{array}$$

Espacial: Lineal. Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre elementos. d_{ij} es la distancia euclídea estimada entre la medida i^{th} y j^{th} , y 1_{ij} es una función de indicador que es 1 si $pd_{ij} \leq 0$ y 0 en caso contrario.

$$\begin{array}{cccc}
(\sigma^2) & \sigma^2(1 - \rho d_{12}) 1_{12} & \sigma^2(1 - \rho d_{13}) 1_{13} & \sigma^2(1 - \rho d_{14}) 1_{14} \\
(\sigma^2(1 - \rho d_{12}) 1_{12}) & \sigma^2 & \sigma^2(1 - \rho d_{23}) 1_{23} & \sigma^2(1 - \rho d_{24}) 1_{24} \\
(\sigma^2(1 - \rho d_{13}) 1_{13}) & \sigma^2(1 - \rho d_{23}) 1_{23} & \sigma^2 & \sigma^2(1 - \rho d_{34}) 1_{34} \\
(\sigma^2(1 - \rho d_{14}) 1_{14}) & \sigma^2(1 - \rho d_{24}) 1_{24} & \sigma^2(1 - \rho d_{34}) 1_{34} & \sigma^2
\end{array}$$

Espacial: Log lineal. Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre elementos. d_{ij} es la distancia euclídea estimada entre la medición de i^{th} y j^{th} , y 1_{ij} es una función de indicador que es 1 si $\rho \log(d_{ij}) \leq 0$ y 0 en caso contrario.

$$\begin{array}{cccc}
(\sigma^2) & \sigma^2(1 - \rho \log(d_{12})) 1_{12} & \sigma^2(1 - \rho \log(d_{13})) 1_{13} & \sigma^2(1 - \rho \log(d_{14})) 1_{14} \\
(\sigma^2(1 - \rho \log(d_{12})) 1_{12}) & \sigma^2 & \sigma^2(1 - \rho \log(d_{23})) 1_{23} & \sigma^2(1 - \rho \log(d_{24})) 1_{24} \\
(\sigma^2(1 - \rho \log(d_{13})) 1_{13}) & \sigma^2(1 - \rho \log(d_{23})) 1_{23} & \sigma^2 & \sigma^2(1 - \rho \log(d_{34})) 1_{34} \\
(\sigma^2(1 - \rho \log(d_{14})) 1_{14}) & \sigma^2(1 - \rho \log(d_{24})) 1_{24} & \sigma^2(1 - \rho \log(d_{34})) 1_{34} & \sigma^2
\end{array}$$

Espacial: Esférico. Esta estructura de covarianza tiene varianzas homogéneas y correlaciones heterogéneas entre elementos. $r_{ij} = d_{ij}/\rho$, donde d_{ij} es la distancia euclídea estimada entre la medición de i^{th} y j^{th} . 1_{ij} es una función de indicador que es 1 si $d_{ij} \leq \rho$ y 0 en caso contrario.

$$\begin{array}{cccc}
(\sigma^2) & \sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} \\
(\sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12}) & \sigma^2 & \sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} \\
(\sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13}) & \sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2 & \sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} \\
(\sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14}) & \sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} & \sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} & \sigma^2
\end{array}$$

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.

$$\begin{array}{cccc}
(\sigma^2) & \sigma^2\rho_1 & \sigma^2\rho_2 & \sigma^2\rho_3 \\
(\sigma^2\rho_1) & \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 \\
(\sigma^2\rho_2) & \sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 \\
(\sigma^2\rho_3) & \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2
\end{array}$$

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.

$$\begin{array}{cccc}
(\sigma_1^2) & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3
\end{array}$$

$(\sigma_2\sigma_1\rho_1$	σ_2^2	$\sigma_3\sigma_2\rho_1$	$\sigma_4\sigma_2\rho_2)$
$(\sigma_3\sigma_1\rho_2$	$\sigma_3\sigma_2\rho_1$	σ_3^2	$\sigma_4\sigma_3\rho_1)$
$(\sigma_4\sigma_1\rho_3$	$\sigma_4\sigma_2\rho_2$	$\sigma_4\sigma_3\rho_1$	$\sigma_4^2)$

Sin estructura. Es una matriz de covarianzas completamente general.

$(\sigma_1^2$	σ_{21}	σ_{31}	$\sigma_{41})$
$(\sigma_{21}$	σ_2^2	σ_{32}	$\sigma_{42})$
$(\sigma_{31}$	σ_{32}	σ_3^2	$\sigma_{43})$
$(\sigma_{41}$	σ_{42}	σ_{43}	$\sigma_4^2)$

Sin estructura: Métrica de correlación. Esta estructura de covarianza tiene varianzas heterogéneas y correlaciones heterogéneas.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_{21}$	$\sigma_3\sigma_1\rho_{31}$	$\sigma_4\sigma_1\rho_{41})$
$(\sigma_2\sigma_1\rho_{21}$	σ_2^2	$\sigma_3\sigma_2\rho_{32}$	$\sigma_4\sigma_2\rho_{42})$
$(\sigma_3\sigma_1\rho_{31}$	$\sigma_3\sigma_2\rho_{32}$	σ_3^2	$\sigma_4\sigma_3\rho_{43})$
$(\sigma_4\sigma_1\rho_{41}$	$\sigma_4\sigma_2\rho_{42}$	$\sigma_4\sigma_3\rho_{43}$	$\sigma_4^2)$

Componentes de la varianza. Esta estructura asigna una estructura de identidad escalada (ID) a cada uno de los efectos aleatorios especificados.

Estadísticas Bayesianas

IBM SPSS Statistics proporciona soporte para las siguientes estadísticas bayesianas.

Pruebas t de muestra única y par de muestras

El procedimiento de inferencia de una muestra bayesiana proporciona opciones para realizar una inferencia bayesiana sobre la prueba t emparejada de una muestra y dos muestras caracterizando distribuciones posteriores. Si tiene datos normales, puede utilizar una previa normal para obtener una posterior normal.

Pruebas de proporción binomial

El procedimiento Inferencia de una muestra bayesiana: binomial proporciona opciones para ejecutar la inferencia de una muestra bayesiana en una distribución binomial. El parámetro de interés es π , que denota la probabilidad de éxito en un número fijo de ensayos que pueden conducir a un éxito o un fracaso. Tenga en cuenta que cada ensayo es independiente, y la probabilidad π sigue siendo la misma en cada ensayo. Una variable aleatoria binomial se puede ver como la suma de un número fijo de ensayos de Bernoulli independientes.

Análisis de distribución de Poisson

El procedimiento Inferencia de una muestra bayesiana: Poisson proporciona opciones para ejecutar la inferencia de una muestra bayesiana sobre la distribución de Poisson. La distribución de Poisson, un modelo útil para eventos raros, supone que dentro de intervalos de tiempo pequeños, la probabilidad de que se produzca un evento es proporcional a la duración del tiempo de espera. Se utiliza un previo de conjugado con la familia de distribución gamma al representar una inferencia estadística bayesiana en una distribución de Poisson.

Muestras relacionadas

El diseño de la inferencia de muestra relacionada bayesiana es bastante similar al de la inferencia de una muestra bayesiana en términos de manejo de las muestras emparejadas. Puede especificar los nombres de variable en pares y ejecutar el análisis Bayesiano en la diferencia de medias.

Pruebas T de muestras independientes

El procedimiento de inferencia de muestra independiente bayesiana proporciona opciones para utilizar una variable de grupo para definir dos grupos no relacionados y realizar una inferencia

bayesiana sobre la diferencia de las dos medias de grupo. Puede estimar los factores bayesianos utilizando distintos enfoques y también caracterizar la distribución posterior deseada ya sea suponiendo que las varianzas sean conocidas o desconocidas.

Correlación por pares (Pearson)

La inferencia bayesiana sobre el coeficiente de correlación de Pearson mide la relación lineal entre dos variables de escala conjuntamente tras una distribución normal bivariada. La inferencia estadística convencional sobre el coeficiente de correlación ha sido ampliamente discutida y su práctica se ha ofrecido durante mucho tiempo en IBM SPSS Statistics. El diseño de la inferencia bayesiana sobre el coeficiente de correlación de Pearson le permite dibujar una inferencia bayesiana estimando factores bayesianos y caracterizando distribuciones posteriores.

Regresión lineal

La inferencia bayesiana sobre la regresión lineal es un método estadístico que se utiliza ampliamente en modelos cuantitativos. La regresión lineal es un enfoque básico y estándar en el cual los investigadores utilizan los valores de varias variables para explicar o predecir valores de un resultado de escala. La regresión lineal univariada Bayesiana es un enfoque a la regresión lineal donde se realiza el análisis estadístico dentro del contexto de la inferencia Bayesiana.

ANOVA unidireccional

El procedimiento ANOVA unidireccional bayesiana genera un análisis de varianza unidireccional para una variable dependiente cuantitativa respecto a una única variable de factor (independiente). El análisis de varianza se utiliza para contrastar la hipótesis de que varias medias son iguales. SPSS Statistics ofrece soporte a factores Bayes, previas de conjugación y previas no informativas.

Modelos de regresión loglineal

El diseño para probar la independencia de dos factores requiere dos variables categóricas para la construcción de una tabla de contingencia, y realiza una inferencia bayesiana sobre la asociación de fila y columna. Puede estimar los factores bayesianos asumiendo distintos modelos, y caracterizar la distribución posterior deseada simulando el intervalo creíble simultáneo para los términos de interacción.

ANOVA de medidas repetidas unidireccional

El procedimiento ANOVA de medidas repetidas unidireccional bayesiano mide un factor del mismo sujeto en cada punto del tiempo o condición distinto, y permite cruzar los sujetos dentro de los niveles. Se da por supuesto que cada sujeto tiene una sola observación para cada punto de tiempo o condición (como tal, la interacción de sujeto-tratamiento no se contabiliza).

Inferencia de una muestra Bayesiana: Normal

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El procedimiento Inferencia de una muestra Bayesiana: Normal proporciona opciones para realizar una inferencia Bayesiana sobre una prueba t emparejada de una muestra y dos muestras caracterizando distribuciones posteriores. Si tiene datos normales, puede utilizar una previa normal para obtener una posterior normal.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Una muestra normal

2. Seleccione las **VARIABLES DE PRUEBA** apropiadas en la lista **VARIABLES DISPONIBLES**. Es necesario seleccionar como mínimo una variable.

Nota: La lista de variables disponibles proporciona todas las variables excepto para las variables de Fecha y Cadena.

3. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.

- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 2. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

1
2

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.
4. Seleccione y/o especifique los valores apropiados de **Varianza de datos** y **valores de hipótesis**. La tabla refleja las variables que están actualmente en la lista **Variables de prueba**. A medida que se añaden o se eliminan variables de la lista **Variables de prueba**, la tabla añade o elimina automáticamente las mismas variables de sus columnas de variables.
- Cuando una o más variables están en la lista **Variables de prueba**, las columnas **Variable conocida** y **Valor de varianza** están habilitadas.
 - Varianza conocida**
Seleccione esta opción para cada variable cuando se conoce la varianza.
 - Valor de varianza**
Un parámetro opcional que especifica el valor de varianza, si se conoce, para datos observados.
 - Cuando una o más variables están en la lista **Variables de prueba** y **Caracterizar distribución posterior** no está seleccionado, las columnas habilitadas son **Valor de prueba nulo** y **Valor g**.
 - Valor de prueba nulo**
Un parámetro necesario que especifica el valor nulo en la estimación de factor Bayesiano. Solo está permitido un valor y 0 es el valor predeterminado.
 - Valor g**
Especifica el valor para definir $\psi^2 = g\sigma^2_x$ en la estimación de factor Bayes. Cuando se especifica el **Valor de varianza**, el valor predeterminado de **Valor g** es 1. Cuando no se especifica el **Valor de varianza**, puede especificar un valor de g fijo u omitir el valor para integrarlo.

¹ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

² Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

5. Si lo desea, puede pulsar **Criterios** para especificar valores de “Inferencia de una muestra Bayesiana: Criterios” en la página 94 (porcentaje de intervalo creíble, opciones de valores perdidos y valores de método numérico), o pulsar **A priori** para especificar valores de “Inferencia de una muestra bayesiana: valores previos normales” en la página 95 (tipo de previos como, por ejemplo, parámetros de inferencia, varianza dada de media o precisión).

Inferencia de una muestra Bayesiana: Criterios

Puede especificar los criterios de análisis siguientes para su inferencia de una muestra Bayesiana:

% Porcentaje de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95 %.

Valores perdidos

Especifique el método en el cual controlar valores perdidos.

Excluir casos según pareja

Este es el valor predeterminado y excluye registros con valores perdidos según análisis. Los registros que incluyen valores perdidos, para un campo que se utiliza para una prueba específica, se omiten de la prueba.

Excluir casos según lista

Este valor excluye registros que incluyen órdenes de lista de valores perdidos. Los registros que incluyen valores perdidos para cualquier campo especificado en cualquier subcomando se excluyen de todos los análisis.

Nota: Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayes** o **Utilizar ambos métodos** para el **Análisis Bayesiano**.

Método numérico

Especifique el método numérico que se utiliza para estimar la integral.

Cuadratura de Gauss-Lobatto adaptativa

Este es el valor predeterminado y llama al enfoque de Cuadratura de Gauss-Lobatto adaptativa.

Tolerancia

Especifique el valor de tolerancia para los métodos numéricos. El valor predeterminado es 0,000001. La opción solo está disponible cuando está seleccionado el valor **Cuadratura de Gauss-Lobatto adaptativa**.

Máximo de iteraciones

Especifique el número máximo de iteraciones de método de la Cuadratura de Gauss-Lobatto adaptativa. El valor debe ser un entero positivo. El valor predeterminado es 2000. La opción solo está disponible cuando está seleccionado el valor **Cuadratura de Gauss-Lobatto adaptativa**.

Aproximación de Monte Carlo

Esta opción llama al enfoque de la aproximación de Monte Carlo.

Establecer semilla personalizada

Cuando está seleccionado, puede especificar un valor de semilla personalizada en el campo **Semilla**.

Semilla

Especifique una semilla aleatoria para el método de aproximación de Monte Carlo. El valor debe ser un entero positivo. De forma predeterminada, se asigna un valor de semilla aleatoria.

Número de muestras Montecarlo

Especifique el número de puntos que se han muestreado para la aproximación Montecarlo. El valor debe ser un entero positivo. El valor predeterminado es 1000000. La opción sólo está disponible cuando el valor **Aproximación de Monte Carlo** ese ha seleccionado.

Inferencia de una muestra bayesiana: valores previos normales

Puede especificar los siguientes criterios de distribución previa para la Inferencia de una muestra bayesiana:

Nota: Muchos investigadores pueden cuestionarse la necesidad de especificar un valor previo. Los valores previos de referencia minimizan la preocupación que surge cuando se desborda el valor previo a medida que aumentan los datos. Cuando se especifica información previa informativa, los métodos bayesianos pueden utilizar de forma eficaz la información. El requisito de especificar una prioridad no debe considerarse como un impedimento para utilizar el análisis bayesiano.

Valor previo en varianza/precisión

Proporciona opciones para definir valores de varianza y precisión.

Varianza

Seleccione para especificar la distribución previa para el parámetro de la varianza. Cuando esta opción está seleccionada, la lista de **Distribución a priori** proporciona las opciones siguientes:

Nota: Si ya se ha especificado la varianza de datos para algunas variables, se ignoran los valores siguientes para estas variables.

- **Difuso** - el valor predeterminado. Especifica el valor previo difuso.
- **Chi cuadrado inverso:** especifica la distribución y los parámetros para $\sigma^2(v_0, \sigma^2_0)$ inverso, donde $\sigma_0 > 0$ es el grado de libertad y $\sigma^2_0 > 0$ es el parámetro de escala.
- **Gamma inverso:** especifica la distribución y los parámetros para gamma inverso (α_0, β_0) , donde $\alpha_0 > 0$ es el parámetro de forma, y $\alpha_0 > 0$ es el parámetro de escala.
- **Jefferys S2:** especifica el valor previo no informativo $\propto 1/\sigma^2_0$.
- **Jefferys S4:** especifica el valor previo no informativo $\propto 1/\sigma^4_0$.

Precisión

Seleccione esta opción para especificar la distribución previa para el parámetro de precisión. Cuando esta opción está seleccionada, la lista de **Distribución a priori** proporciona las opciones siguientes:

- **Gamma:** especifica la distribución y parámetros para gamma (α_0, β_0) , donde $\alpha_0 > 0$ es el parámetro de forma, y $\alpha_0 > 0$ es el parámetro de escala.
- **Chi cuadrado:** especifica la distribución y parámetros para $\alpha^2(v_0)$, donde $\alpha_0 > 0$ es el grado de libertad.

Parámetro de forma

Especifique el parámetro de forma a_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0.

Parámetro de escala

Especifique el parámetro de escala b_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0. Cuanto mayor sea el parámetro de escala, más dispersa será la distribución.

Valor previo sobre una media dada la varianza/precisión

Especifique la distribución previa para el parámetro de media que es condicional a la varianza o al parámetro de precisión.

Normal

Especifica la distribución y los parámetros para Normal $(\mu_0, K^{-1}_0\sigma^2_0)$ en la varianza o Normal $(\mu_0, K_0/\sigma^2_0)$ en la precisión, donde $\mu_0 \in (-\infty, \infty)$ and $\sigma^2 > 0$.

Parámetro de ubicación

Especifique un valor numérico que especifica el parámetro de ubicación para la distribución.

Parámetro de escala

Especifique el parámetro de escala b_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0.

Kappa

Especifique el valor de K_0 en $\text{Normal}(\mu_0, K_0^{-1}\sigma_0^2)$ o $\text{Normal}(\mu_0, K_0/\sigma_0^2)$. Debe escribir un valor único que sea mayor que 0 (1 es el valor predeterminado).

Difuso

El valor predeterminado que especifica el valor previo difuso $\propto 1$.

Inferencia de una muestra Bayesiana: Binomial

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El procedimiento Inferencia de una muestra bayesiana: binomial proporciona opciones para ejecutar la inferencia de una muestra bayesiana en una distribución binomial. El parámetro de interés es π , que denota la probabilidad de éxito en un número fijo de ensayos que pueden conducir a un éxito o un fracaso. Tenga en cuenta que cada ensayo es independiente, y la probabilidad π sigue siendo la misma en cada ensayo. Una variable aleatoria binomial puede considerarse la suma de un número fijo de ensayos independientes Bernoulli.

Aunque no es necesario, normalmente se elige una previa de la familia de distribución Beta al estimar un parámetro binomial. La familia Beta es un conjugado para la familia binomial y, como tal, conduce a la distribución posterior con un formulario cerrado aún en la familia de distribución Beta.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Binomial de una muestra

2. Seleccione las **Variables de prueba** apropiadas en la lista **Variables disponibles**. Es necesario seleccionar como mínimo una variable.

Nota: La lista de variables disponibles proporciona todas las variables excepto para las variables de Fecha y Cadena.

3. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1

Tabla 3. Umbrales utilizados comúnmente para definir la significación de las pruebas (continuación)

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

3

4

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.
4. Seleccione y/o escriba los valores **Categorías de éxito y valores de hipótesis**. La tabla refleja las variables que están actualmente en la lista **Variables de prueba**. A medida que se añaden o eliminan variables de la lista **Variables de prueba**, la tabla añade o suprime automáticamente las mismas variables de sus columnas de pares de variables.
- Cuando **Caracterizar distribución posterior** se selecciona como el **Análisis Bayesiano**, la columna **Categorías de éxito** se habilitan.
 - Cuando se selecciona **Estimación del factor Bayes** o **Utilizar ambos métodos** como el **Análisis Bayesiano**, todas las columnas editables se habilitan.

Punto nulo

Habilita e inhabilita la opción **Proporción nula**. Cuando el valor está habilitado, ambas opciones **Forma previa nula** y **Escala previa nula** están inhabilitadas.

Forma anterior de nulo

Especifica el parámetro de forma a_0 bajo la hipótesis nula de inferencia binomial.

Escala previa de nulo

Especifica el parámetro de escala b_0 bajo la hipótesis nula de inferencia binomial.

Proporción de nulos

Especifica el parámetro de forma a_0 y el parámetro de escala b_0 bajo la hipótesis nula para una distribución previa de conjugación (para acomodar las previas de Beta y Haldane). El rango válido son valores numéricos entre 0 y 1.

Forma anterior alternativa

Un parámetro necesario para especificar a_0 bajo la hipótesis alternativa de inferencia binomial si se va a estimar el factor Bayes.

Escala previa alternativa

Un parámetro necesario para especificar b_0 bajo la hipótesis alternativa de inferencia binomial si se va a estimar el factor Bayes.

Categorías de éxito

Proporciona opciones para definir distribuciones previas de conjugación. Las opciones proporcionadas especifican cómo se define el éxito, cuando los valores de datos se prueban con respecto al valor de prueba.

Última categoría

El valor predeterminado que realiza la prueba binomial utilizando el último valor numérico encontrado en la categoría después de que se haya clasificado en un orden ascendente.

³ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

⁴ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Primera categoría

Realiza la prueba binomial utilizando el primer valor numérico encontrado en la categoría después de almacenarse en un orden ascendente.

Punto medio

Utiliza el punto medio de los valores numéricos como casos. Un punto medio es el promedio de los datos de muestra mínimo y máximo.

Punto de corte

Utiliza un valor de corte especificado de valores numéricos como casos. El valor debe ser un valor numérico único.

Nivel

Trata los valores de cadena especificados por el usuario (puede ser más de 1) como casos. Utilice comas para separar los distintos valores.

5. Si lo desea puede pulsar **Criterios** para especificar valores de “Inferencia de una muestra Bayesiana: Criterios” en la página 94 (porcentaje de intervalo creíble, opciones de valores perdidos y valores de método numérico), o pulse **A priori** para especificar valores de “Inferencia de muestra única Bayesiana: A priori binomiales/Poisson” en la página 98 (conjugado o distribuciones de previas personalizadas).

Inferencia de muestra única Bayesiana: A priori binomiales/Poisson

Puede especificar los siguientes criterios de distribución previa para la Inferencia de una muestra bayesiana:

Nota: Muchos investigadores pueden cuestionarse la necesidad de especificar un valor previo. Los valores previos de referencia minimizan la preocupación que surge cuando se desborda el valor previo a medida que aumentan los datos. Cuando se especifica información previa informativa, los métodos bayesianos pueden utilizar de forma eficaz la información. El requisito de especificar una prioridad no debe considerarse como un impedimento para utilizar el análisis bayesiano.

Parámetro de forma

Para previas binomiales, especifique el parámetro de forma α_0 para la distribución Beta.

Para previas Poisson, especifique el parámetro de forma α_0 para la distribución gamma.

Debe especificar un único valor que sea mayor que 0.

Parámetro de escala

Para previas binomiales, especifique el parámetro de escala b_0 para la distribución Beta.

Para previas Poisson, especifique el parámetro de escala b_0 para la distribución gamma.

Debe especificar un único valor que sea mayor que 0.

Inferencia de una muestra Bayesiana: Poisson

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El procedimiento Inferencia de una muestra bayesiana: Poisson proporciona opciones para ejecutar la inferencia de una muestra bayesiana sobre la distribución de Poisson. La distribución de Poisson, un modelo útil para eventos raros, supone que dentro de intervalos de tiempo pequeños, la probabilidad de que se produzca un evento es proporcional a la duración del tiempo de espera. Se utiliza un previo de conjugado con la familia de distribución gamma al representar una inferencia estadística bayesiana en una distribución de Poisson.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Poisson de una muestra

2. Seleccione las **Variables de prueba** apropiadas en la lista **Variables disponibles**. Es necesario seleccionar como mínimo una variable.

Nota: La lista de variables disponibles proporciona todas las variables excepto para las variables de Fecha y Cadena.

3. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 4. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

5

6

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.
4. Seleccione y/o especifique los valores apropiados de **Valores de hipótesis**. La tabla refleja las variables que están actualmente en la lista **Variables de prueba**. Puesto que se añaden o eliminan variables de **Variables de prueba**, la tabla añade o elimina automáticamente las mismas variables de sus columnas de par de variables.

- Cuando está seleccionado **Caracterizar distribución a priori** como el **Análisis Bayesiano**, ninguna de las columnas está habilitada.
- Cuando **Estimar factor Bayes** o **Utilizar ambos métodos** están seleccionados como el **Análisis Bayesiano**, están habilitadas todas las columnas editables.

Punto nulo

Habilita o inhabilita la opción **Índice nulo**. Cuando el valor está habilitado, ambas opciones **Forma previa nula** y **Escala previa nula** están inhabilitadas.

Forma anterior de nulo

Especifica el parámetro de forma α_0 debajo de la hipótesis nula de inferencia Poisson.

⁵ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

⁶ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Escala previa de nulo

Especifica el parámetro de escala b_0 debajo de la hipótesis nula de inferencia Poisson.

Índice de nulos

Especifica el parámetro de forma a_0 y el parámetro de escala b_0 bajo la hipótesis nula para una distribución previa de conjugación (para acomodar las previas de Poisson-Gamma). El valor mínimo debe ser un valor numérico de valor superior a 0; el valor máximo debe ser un valor doble máximo.

Forma anterior alternativa

Un parámetro necesario para especificar a_1 debajo de la hipótesis alternativa de inferencia Poisson si se va a estimar el factor Bayesiano.

Escala previa alternativa

Un parámetro necesario para especificar b_1 debajo de la hipótesis alternativa de inferencia Poisson si se va a estimar el factor Bayesiano.

5. Si lo desea puede pulsar **Criterios** para especificar valores de “Inferencia de una muestra Bayesiana: Criterios” en la página 94 (porcentaje de intervalo creíble, opciones de valores perdidos y valores de método numérico), o pulse **A priori** para especificar valores de “Inferencia de muestra única Bayesiana: A priori binomiales/Poisson” en la página 98 (conjugado o distribuciones de previas personalizadas).

Inferencia de muestra relacionada Bayesiana: Normal

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El procedimiento Inferencia de muestra relacionada Bayesiana: Normal proporciona opciones de inferencia de una muestra Bayesiana para muestras emparejadas. Puede especificar los nombres de variables en pares y ejecutar el análisis Bayesiano en la diferencia de medias.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Muestras relacionadas normales

2. Seleccione las **Variables emparejadas** apropiadas en la lista **Variables disponibles**. Como mínimo, se debe seleccionar un par de variables de origen y no se pueden seleccionar más de dos variables de origen para cualquier conjunto de pares determinado.

Nota: La lista de variables disponibles proporciona todas las variables excepto para las variables de cadena.

3. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 5. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1

Tabla 5. Umbrales utilizados comúnmente para definir la significación de las pruebas (continuación)

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

7

8

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.
4. Seleccione y/o especifique los valores apropiados de **Varianza de datos y valores de hipótesis**. La tabla refleja los pares de variables que están actualmente en la lista **Variables emparejadas**. A medida que se añaden o se eliminan pares de variables en la lista **Variables emparejadas**, la tabla añade o elimina automáticamente los mismos pares de variables de las columnas de pares de variables.
- Cuando uno o más pares de variables están en la lista **Variables emparejadas**, las columnas **Varianza conocida** y **Valor de varianza** están habilitadas.
- Varianza conocida**
 Seleccione esta opción para cada variable cuando se conoce la varianza.
- Valor de varianza**
 Un parámetro opcional que especifica el valor de varianza, si se conoce, para datos observados.
- Cuando hay uno o varios pares de variables en la lista **Variables emparejadas** y **Caracterizar distribución posterior** no está seleccionado, se habilitan las columnas **Valor de prueba nulo** y **Valor g**.
- Valor de prueba nulo**
 Un parámetro necesario que especifica el valor nulo en la estimación de factor Bayesiano. Solo está permitido un valor y 0 es el valor predeterminado.
- Valor g**
 Especifica el valor para definir $\psi^2 = g\sigma^2_x$ en la estimación de factor Bayes. Cuando se especifica el **Valor de varianza**, el valor predeterminado de **Valor g** es 1. Cuando no se especifica el **Valor de varianza**, puede especificar un valor de g fijo u omitir el valor para integrarlo.
5. Si lo desea puede pulsar **Criterios** para especificar valores de “Inferencia de una muestra Bayesiana: Criterios” en la página 94 (porcentaje de intervalo creíble, opciones de valores perdidos y valores de método numérico), o pulse **A priori** para especificar valores de “Inferencia de muestra única Bayesiana: A priori binomiales/Poisson” en la página 98 (conjugado o distribuciones de previas personalizadas).

Inferencia de muestra independiente Bayesiana

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

⁷ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

⁸ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

El procedimiento Inferencia de muestra independiente Bayesiana proporciona opciones para utilizar una variable de grupo para definir dos grupos no relacionados y realizar una inferencia Bayesiana sobre la diferencia de las dos medias de grupo. Puede estimar los factores bayesianos utilizando distintos enfoques y también caracterizar la distribución posterior deseada ya sea suponiendo que las varianzas sean conocidas o desconocidas.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Muestras independientes normales

2. Seleccione las **Variables de prueba** adecuadas de la lista de variables de origen. Al menos, se debe seleccionar una variable de origen.
3. Seleccione la **Variable de agrupación** apropiada en la lista **Variables disponibles**. Una variable de agrupación define dos grupos para la prueba *t* no emparejada. La variable de agrupación seleccionada puede ser una variable numérica o de serie.
4. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 6. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

9

10

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.

⁹ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁰ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

5. Utilice las opciones **Definir grupos** para definir dos grupos para la prueba t especificando dos valores (para variables de cadena), o dos valores, un punto medio, o un punto de corte (para variables numéricas).

Nota: Los valores especificados deben estar en la variable, de lo contrario se muestra un mensaje de error para indicar que al menos uno de los grupos está vacío.

Para variables numéricas

- **Usar valores especificados.** Especifique un valor para el grupo 1 y otro valor para el grupo 2. Los casos con cualquier otro valor se excluyen del análisis. Los números no tienen que ser enteros (por ejemplo, 6,25 y 12,5 son válidos).
- **Utilice el valor de punto medio.** Cuando está seleccionado, los grupos se separan en $<$ y valores de punto medio.
- **Utilizar punto de corte.**
 - **Punto de corte.** Escriba un número que divida los valores de la variable de agrupación en dos conjuntos. Todos los casos con valores menores que el punto de corte forman un grupo y los casos con valores mayores o iguales que el punto de corte forman el otro grupo.

Para las variables de agrupación de series, especifique una serie para el grupo 1 y otro valor para el grupo 2, como por ejemplo *Sí* y *no*. Los casos con otras series se excluyen del análisis.

6. De forma opcional, puede pulsar **Criterios** para especificar valores de “Inferencia de muestra independiente Bayesiana: Criterios” en la página 103 (porcentaje de intervalo creíble, opciones de valores perdidos y valores de método de cuadratura adaptativa), pulse **A priori** para especificar valores de “Inferencia de muestra independiente Bayesiana: Distribución a priori” en la página 104 (varianza de datos, a priori de varianza y a priori en condicional en varianza), o pulse **Estimar Factor Bayesiano** para especificar valores de “Independiente Bayesiano - Inferencia de muestra: Estimación del factor Bayes” en la página 105.

Inferencia de muestra independiente Bayesiana - Definir grupos (numérico)

Para las variables de agrupación numérica, defina los dos grupos para la prueba t especificando dos valores, un punto medio o un punto de corte.

Nota: Los valores especificados deben estar en la variable, de lo contrario se muestra un mensaje de error para indicar que al menos uno de los grupos está vacío.

- **Usar valores especificados.** Especifique un valor para el grupo 1 y otro valor para el grupo 2. Los casos con cualquier otro valor se excluyen del análisis. Los números no tienen que ser enteros (por ejemplo, 6,25 y 12,5 son válidos).
- **Utilice el valor de punto medio.** Cuando está seleccionado, los grupos se separan en $<$ y valores de punto medio.
- **Utilizar punto de corte.**
 - **Punto de corte.** Escriba un número que divida los valores de la variable de agrupación en dos conjuntos. Todos los casos con valores menores que el punto de corte forman un grupo y los casos con valores mayores o iguales que el punto de corte forman el otro grupo.

Inferencia de muestra independiente Bayesiana - Definir grupos (cadena)

Para las variables de agrupación de series, especifique una serie para el grupo 1 y otro valor para el grupo 2, como por ejemplo *Sí* y *no*. Los casos con otras series se excluyen del análisis.

Nota: Los valores especificados deben estar en la variable, de lo contrario se muestra un mensaje de error para indicar que al menos uno de los grupos está vacío.

Inferencia de muestra independiente Bayesiana: Criterios

Puede especificar los siguientes criterios de análisis para la inferencia de una muestra Bayesiana independiente:

Porcentaje % de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95%.

Valores perdidos

Especifique el método con el que controlar los valores perdidos.

Excluir casos según pareja

Este es el valor predeterminado y excluye registros con valores perdidos según análisis. Los registros que incluyen valores perdidos, para un campo que se utiliza para una prueba específica, se omiten de la prueba.

Excluir casos según lista

Este valor excluye registros que incluyen órdenes de lista de valores perdidos. Los registros que incluyen valores perdidos para cualquier campo especificado en cualquier subcomando se excluyen de todos los análisis.

Nota: Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayes** o **Utilizar ambos métodos** para el **Análisis Bayesiano**.

Método de cuadratura adaptativa

Especifique los valores de iteración máximo y de tolerancia para la cuadratura adaptativa.

Tolerancia

Especifique el valor de tolerancia para los métodos numéricos. El valor predeterminado es 0,000001.

Número máximo de iteraciones

Especifique el número máximo de iteraciones de método de cuadratura adaptativa. El valor debe ser un entero positivo. El valor predeterminado es 500.

Inferencia de muestra independiente Bayesiana: Distribución a priori

Puede especificar los siguientes criterios de distribución previa para la Inferencia de muestra independiente Bayesiana:

Nota: Muchos investigadores pueden cuestionarse la necesidad de especificar un valor previo. Los valores previos de referencia minimizan la preocupación que surge cuando se desborda el valor previo a medida que aumentan los datos. Cuando se especifica información previa informativa, los métodos bayesianos pueden utilizar de forma eficaz la información. El requisito de especificar una prioridad no debe considerarse como un impedimento para utilizar el análisis bayesiano.

Varianza de datos

Proporciona opciones para definir valores de varianza de datos.

Varianza conocida

Cuando selecciona esta opción, le permite especificar dos varianzas de grupo conocidas. Ambos valores deben ser > 0 .

Varianza de grupo 1

Escriba el primer valor de varianza de grupo conocido.

Varianza de grupo 2

Escriba el segundo valor de varianza de grupo conocido.

Suponer varianza igual

Controla si se supone o no que las dos varianzas de grupo son iguales. De forma predeterminada, se da por supuesto que las varianzas de grupo no son iguales. Este valor se ignora cuando se escriben valores para las dos varianzas de grupo.

Suponer varianza desigual

Controla si se da por supuesto que las dos varianzas de grupo van a ser desiguales o no. De forma predeterminada, se da por supuesto que las varianzas de grupo no son iguales. Este valor se ignora cuando se escriben valores para las dos varianzas de grupo.

Previa en varianza

Especifique la distribución a priori para las dos varianzas iguales.

Jeffreys

Cuando está seleccionado, se utiliza una distribución a priori no informativa (objetiva) para un espacio de parámetro.

Inverso-Chi cuadrado

Especifica la distribución de probabilidad continua de una variable aleatoria con un valor positivo y los parámetros para el inverso- $\chi^2(v_0, \sigma^2_0)$, donde $v_0 > 0$ es el grado de libertad y $\sigma^2_0 > 0$ es el parámetro de escala.

Grados de libertad

Especifique un valor para el número de valores en el cálculo final que pueden variar.

Parámetro de escala

Especifique el parámetro de escala $\sigma^2_0 > 0$ para inverso- $\chi^2(v_0, \sigma^2_0)$. Debe especificar un único valor que sea mayor que 0. Cuanto mayor sea el parámetro de escala, más dispersa será la distribución.

A priori en condicional de media sobre varianza

Proporciona opciones para especificar la distribución a priori para las dos medias de grupo.

Nota: Las opciones **Difuso** y **Normal** sólo están disponibles cuando la opción **Varianza conocida** está seleccionada.

Difuso

El valor predeterminado. Especifica el valor previo difuso.

Normal

Cuando se selecciona esta opción, debe especificar los parámetros de ubicación y escala para las medias de grupo definidas.

Parámetro de ubicación

Entre un valor numérico que especifica los parámetros de ubicación para las distribuciones de grupo.

Parámetro de escala

Especifique el parámetro de escala $\sigma^2_0 > 0$ para inverso- $\chi^2(v_0, \sigma^2_0)$. Para cada grupo, debe especificar un único valor que sea mayor que 0. Cuanto mayor sea el parámetro de escala, más dispersa será la distribución.

Independiente Bayesiano - Inferencia de muestra: Estimación del factor Bayes

Puede especificar el método que se utiliza para estimar el factor Bayes.

Método de Rouder

Cuando se selecciona esta opción, se invoca el método de Rouder. Es el valor predeterminado

Método de Gonen

Cuando está seleccionado, invoca el enfoque Gonen y debe especificar los valores siguientes de tamaño de efecto:

Media de tamaño de efecto

Especifique un valor que indique la diferencia de medias entre los dos grupos.

Varianza del tamaño de efecto

Especifique un valor que indique la varianza para los dos grupos. El valor debe ser > 0 .

Método hyper previo

Cuando se selecciona, invoca el enfoque hiper-g cuando debe especificar un solo valor. Especifique un valor entre -1 y -0,5 en el campo **Parámetro de forma**. El valor predeterminado es 0,75.

Inferencia Bayesiana sobre la correlación de Pearson

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El coeficiente de correlación de Pearson mide la relación lineal entre dos variables de escala conjuntamente después de una distribución normal bivalente. La inferencia estadística convencional sobre el coeficiente de correlación se ha tratado ampliamente y su práctica se ha ofrecido durante mucho tiempo en IBM SPSS Statistics. El diseño de la inferencia Bayesiana sobre el coeficiente de correlación de Pearson permite a los usuarios trazar una inferencia Bayesiana estimando factores Bayesianos y caracterizando distribuciones a priori.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Correlación de Pearson

2. Seleccione las **Variables de prueba** adecuadas para utilizar para la inferencia de correlación por parejas en la lista **Variables disponibles**. Al menos, se deben seleccionar dos variables de origen. Cuando están seleccionadas más de dos variables, el análisis se ejecuta en todas las combinaciones por parejas de las variables seleccionadas.

3. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 7. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

11

12

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.

4. Especifique el **Número máximo de gráficos** para ver en la salida. Un conjunto de gráficos puede contener 3 gráficos en el mismo panel. Los gráficos se generan en orden desde la primera variable

¹¹ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹² Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

versus las variables restantes, después, la segunda variable versus las variables restantes y, así, sucesivamente. El valor entero definido debe estar entre 0 y 50. De forma predeterminada, se generan 10 conjuntos de gráficos para dar cabida a cinco variables. Esta opción no está disponible cuando está seleccionado **Estimar factor Bayesiano**.

5. Opcionalmente puede pulsar **Criterios** para especificar los valores de “Correlación de Pearson Bayesiana: Criterios” en la página 107 (porcentaje de intervalo creíble, opciones de valores que faltan y valores de método numérico), pulsar **Previas** para especificar los valores de “Correlación de Pearson Bayesiana: Distribución previa” en la página 107 (valor c para la previa $p(\rho) \propto (1 - \rho^2)^c$ o pulsar **Factor bayesiano** para especificar los valores de “Independiente Bayesiano - Inferencia de muestra: Estimación del factor Bayes” en la página 105.

Correlación de Pearson Bayesiana: Criterios

Puede especificar los siguientes criterios de análisis de Inferencia de correlación de Pearson Bayesiana (por parejas).

Porcentaje % de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95%.

Valores perdidos

Especifique el método con el que controlar los valores perdidos.

Excluir casos según pareja

Este valor excluye registros que incluyen parejas de valores perdidos.

Excluir casos según lista

Este valor excluye registros que incluyen órdenes de lista de valores perdidos. Los registros que incluyen valores perdidos para cualquier campo especificado en cualquier subcomando se excluyen de todos los análisis.

Nota: Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayes** o **Utilizar ambos métodos** para el **Análisis Bayesiano**.

Método numérico

Especifique el método numérico utilizado para estimar la integral.

Establecer semilla personalizada

Cuando está seleccionado, puede especificar un valor de semilla personalizada en el campo **Semilla**.

Tolerancia

Especifique el valor de tolerancia para los métodos numéricos. El valor predeterminado es 0,000001.

Número máximo de iteraciones

Especifique el número máximo de iteraciones del método. El valor debe ser un entero positivo. El valor predeterminado es 2000.

Número de muestras de Monte Carlo

Especifique el número de puntos que se han muestreado para la aproximación Montecarlo. El valor debe ser un entero positivo. El valor predeterminado es 10000.

Muestras simuladas para distribución posterior

Especifique el número de muestras que se utilizan para igualar la distribución a priori deseada. El valor predeterminado es 10000.

Correlación de Pearson Bayesiana: Distribución previa

Puede especificar el valor c para la previa $p(\rho) \propto (1 - \rho^2)^c$.

Nota: Muchos investigadores pueden cuestionarse la necesidad de especificar un valor previo. Los valores previos de referencia minimizan la preocupación que surge cuando se desborda el valor previo a medida que aumentan los datos. Cuando se especifica información previa informativa, los métodos

bayesianos pueden utilizar de forma eficaz la información. El requisito de especificar una prioridad no debe considerarse como un impedimento para utilizar el análisis bayesiano.

Uniforme (c = 0)

Cuando se selecciona se utiliza el uniforme anterior.

Jeffreys (c = -1.5)

Cuando está seleccionado, se utiliza una distribución a priori no informativa.

Establecer valor c personalizado

Cuando se selecciona, puede especificar un **valor c** personalizado. Se permite cualquier número real único.

Correlación de Pearson Bayesiana: Factor Bayesiano

Puede especificar el método que se utiliza para estimar el factor Bayes. Las siguientes opciones sólo están disponibles cuando se selecciona la opción **Estimación del factor Bayes** o **Utilizar ambos métodos** para Análisis Bayesiano.

Factor bayesiano JZS

Cuando está seleccionado, invocar el enfoque de Zellner-Siow. Ésta es el ajuste predeterminado.

Factor Bayes fraccional

Cuando está seleccionado, puede especificar el factor Bayes fraccional y un valor de hipótesis nula. Para el factor Bayes fraccional, debe especificar un valor (0,1). El valor predeterminado es 0,5.

Inferencia Bayesiana sobre modelos de regresión lineal

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

La regresión es un método estadístico que se utiliza ampliamente en modelos cuantitativos. La regresión lineal es un enfoque básico y estándar en el cual los investigadores utilizan los valores de varias variables para explicar o predecir valores de un resultado de escala. La regresión lineal univariada Bayesiana es un enfoque a la regresión lineal donde se realiza el análisis estadístico dentro del contexto de la inferencia Bayesiana.

Puede invocar el procedimiento de regresión y definir un modelo completo.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Regresión lineal

2. Seleccione una sola variable dependiente no de cadena en la lista **Variables disponibles**. Debe seleccionar una variable no de cadena.

3. Seleccione una o más variables de factor categóricas para el modelo desde la lista **Variables disponibles**.

4. Seleccione una o más variables de escala de covariable de no cadena en la lista **Variables disponibles**.

Nota: Ninguna de las listas, **Factor(es)** y **Covariable(s)** puede estar vacía. Debe seleccionar, al menos, una variable **Factor(es)** o **Covariable(s)**.

5. Opcionalmente, seleccione una sola variable de no cadena para que actúe como la ponderación de regresión desde la lista **Variables disponibles**.

6. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 8. Umbral utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

13

14

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.

Si lo desea, puede:

- Pulse **Criterios** para especificar el porcentaje de intervalo creíble y valores de método numérico.
- Pulse **A priori** para definir valores de referencia y de distribución a priori de conjugado.
- Pulsar **Factor Bayes** para especificar los valores de factor de Bayes.
- Pulse **Guardar** para identificar qué elementos guardar, y guardar información de modelo en un archivo XML.
- Pulsar **Pronosticar** para especificar regresores para la predicción Bayesiana.
- Pulse **Gráficos** para trazar las distribuciones posteriores de los parámetros de regresión, la varianza de términos de error y los valores pronosticados.
- Pulse **Pruebas F** para comparar modelos estadísticos para poder identificar el modelo que se ajusta mejor a la población utilizada para el muestreo.

Modelos de regresión lineal Bayesiana: Criterios

Puede especificar los criterios de análisis siguientes para modelos de regresión lineal Bayesiana.

Porcentaje % de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95%.

Nota: Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayes** o **Utilizar ambos métodos** para el **Análisis Bayesiano**.

Método numérico

Especifique el método numérico utilizado para estimar la integral.

¹³ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁴ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Tolerancia

Especifique el valor de tolerancia para los métodos numéricos. El valor predeterminado es 0,000001.

Número máximo de iteraciones

Especifique el número máximo de iteraciones del método. El valor debe ser un entero positivo. El valor predeterminado es 2000.

Modelos de regresión lineal Bayesianos: Distribuciones a priori

Puede especificar los valores de distribución a priori siguientes para los parámetros de regresión y la varianza de los errores. Las opciones siguientes están disponibles solo cuando la opción **Caracterizar distribución a priori** está seleccionada para el **Análisis Bayesiano**.

Nota: Muchos investigadores pueden cuestionarse la necesidad de especificar un valor previo. Los valores previos de referencia minimizan la preocupación que surge cuando se desborda el valor previo a medida que aumentan los datos. Cuando se especifica información previa informativa, los métodos bayesianos pueden utilizar de forma eficaz la información. El requisito de especificar una prioridad no debe considerarse como un impedimento para utilizar el análisis bayesiano.

Referencia

Cuando está seleccionado, el análisis de referencia genera una inferencia Bayesiana objetiva. Las sentencias inferenciales solo dependen del modelo asumido y los datos disponibles y la distribución a priori que se utiliza para realizar una inferencia es la menos informativa. Ésta es el ajuste predeterminado.

Conjugar

Proporciona opciones para definir distribuciones previas de conjugado. El a priori de conjugado da por supuesto la distribución de unión gamma-inversa-normal. Aunque los a priori de conjugado no son necesarios al realizar actualizaciones Bayesianas, ayudan a los procesos de cálculo.

Nota: Para poder especificar previas de conjugado para un modelo de regresión lineal, establezca la media esperada de parámetros de regresión en la tabla **Previas sobre varianza de errores**. También puede elegir utilizar los valores de **Varianza de matriz de covarianzas** para especificar la varianza-covarianza previa.

A priori en varianza de errores**Parámetro de forma**

Especifique el parámetro de forma α_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0.

Parámetro de escala

Especifique el parámetro de escala b_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0. Cuanto mayor sea el parámetro de escala, más dispersa será la distribución.

La tabla lista la media de parámetros de regresión (incluyendo la intersección), que especifica el vector principal θ_0 para los parámetros de regresión definidos. El número de valores debe cumplir el número de parámetros de regresión, incluyendo el término de intersección.

El primer nombre de variable siempre es INTERCEPT. Desde la segunda fila, la columna **Variables** se llena automáticamente con las variables que se especifican mediante Factores y Covariables. La columna **Media** no incluye ningún valor predeterminado.

Pulse **Restablecer** para borrar los valores.

Varianza de matriz de covarianzas: 2x

Especifique los valores V_0 en el triángulo inferior en la matriz de varianza-covarianza para el a priori normal multivariante. Tenga en cuenta que V_0 debe ser un definitivo semipositivo. El último valor de cada fila debe ser positivo. La siguiente fila debe tener un valor más que la fila anterior. No se han especificado valores para categorías de referencia (si hay alguna).

Pulse **Restablecer** para borrar los valores.

Utilizar matriz de identidad

Cuando está seleccionado, se utiliza la matriz de identidad escalada. No puede especificar valores V_0 en el triángulo inferior en la matriz de varianza-covarianza para el a priori normal multivariante.

Modelos de regresión lineal Bayesianos: Factor Bayes

Puede especificar el diseño de modelo para el análisis, incluido el enfoque que se utiliza para calcular el factor Bayes para Modelos de regresión lineal. Las opciones siguientes solo están disponibles cuando está seleccionada la opción de Análisis Bayesiano **Estimar factor Bayesiano** o **Utilizar ambos métodos**.

Modelo nulo

Cuando se selecciona esta opción, los factores Bayes estimados se basan en el modelo nulo. Ésta es el ajuste predeterminado.

Modelo completo

Cuando se selecciona, los factores Bayesianos estimados se basan en el modelo completo y puede seleccionar variables para utilizar y factores y covariables adicionales.

Variables

Lista todas las variables disponibles para el modelo completo.

Factores adicionales

Seleccione variables en la lista **Variables** para utilizar factores adicionales.

Covariables adicionales

Seleccione variables en la lista **Variables** para utilizarlas como covariables adicionales.

Cálculo

Especifique el enfoque para estimar factores Bayesianos. El método JZS es el valor predeterminado.

Método JZS

Cuando está seleccionado, invocar el enfoque de Zellner-Siow. Ésta es el ajuste predeterminado.

Método de Zellner

Cuando se selecciona, invoca el enfoque de Zellner y se le pedirá que especifique un único valor anterior $g > 0$ (no hay ningún valor predeterminado).

Método hyper previo

Cuando se selecciona, invoca el enfoque hiper g y se le pedirá que especifique un parámetro de forma α_0 para la distribución gamma inversa. Debe especificar un único valor > 0 (el valor predeterminado es 3).

Método de Rouder

Cuando se selecciona, invoca el enfoque de Rouder y se le pedirá que especifique un parámetro de escalado b_0 para la distribución gamma inversa. Debe especificar un único valor > 0 (el valor predeterminado es 1).

Modelos de regresión lineal Bayesianos: Guardar

Este diálogo permite especificar qué estadísticas puntúan para la distribución de predicción Bayesiana y exportar los resultados del modelo a un archivo XML.

Estadísticas predictivas posteriores

Puede puntuar las siguientes estadísticas que derivan de predicciones Bayesianas.

Medias

Media de la distribución predictiva posterior.

Varianzas

Varianza de la distribución predictiva posterior.

Modo

Modo de la distribución predictiva posterior.

Límite inferior de intervalo creíble

Límite inferior del intervalo creíble de la distribución predictiva posterior.

Límite superior de intervalo creíble

Límite superior del intervalo creíble de la distribución predictiva posterior.

Nota: Puede asignar nombres de variable correspondientes para cada estadística.

Exportar información del modelo a un archivo XML

Especifique un nombre de archivo XML y ubicación para exportar la matriz de varianza-covarianza del parámetro puntuado.

Modelos de regresión lineal Bayesianos: Predecir

Puede especificar los regresores para generar distribuciones predictivas.

Regresores para la predicción Bayesiana

La tabla lista todos los regresores disponibles. La columna **Regresores** se llena automáticamente mediante variables de factor y covariable. Especifique los vectores observados con los valores para los regresores. A cada regresor se le puede asignar un valor o cadena y se les permite predecir sólo un caso. Para los factores, ambos valores y cadenas están permitidos.

Se deben especificar todos los valores de regresor o ninguno para poder ejecutar la predicción (pulsando **Ejecutar análisis**).

Cuando se elimina una variable de factor o covariable, la fila de regresor correspondiente se elimina de la tabla.

Para las covariables, sólo se pueden especificar valores numéricos. Para los factores, ambos valores numéricos y cadenas están permitidos.

Nota: Pulse **Restablecer** para borrar los valores definidos.

Modelos de regresión lineal Bayesianos: Gráficos

Puede controlar los gráficos que se generan.

Covariables

Lista las covariables definidas actualmente.

Covariables de gráfico

Seleccione las covariables que se van a trazar en la lista **Covariables** y añádalas a la lista **Covariables de gráfico**.

Factores

Lista los factores definidos actualmente.

Factores de gráfico

Seleccione los factores que se van a trazar desde la lista **Factores** y añádalos a la lista **Trazar factor(es)**.

Máximo de categorías que se van a trazar

Seleccione el número máximo de categorías que se va a trazar (entero positivo, único). El valor se aplica a todos los factores. De forma predeterminada, los primeros 2 niveles se trazan para cada factor.

Incluir gráficos de**Término de interceptación**

Cuando está seleccionado, se traza el término de interceptación. El valor está deseleccionado de forma predeterminada.

Varianza de términos de error

Cuando se selecciona, se traza un gráfico de la varianza de errores. El valor está deseleccionado de forma predeterminada.

Distribución predicha bayesiana

Cuando se selecciona esta opción, se traza la distribución predictiva. El valor está deseleccionado de forma predeterminada. El valor sólo se puede seleccionar cuando se seleccionan valores de regresores válidos.

Modelos de regresión lineal Bayesianos: Pruebas-F

Puede crear una o más pruebas F parciales. Una prueba F es una prueba estadística en la que la estadística de la prueba tiene una distribución F bajo la hipótesis nula. Normalmente, las pruebas F se utilizan al comparar modelos estadísticos que se han ajustado a un conjunto de datos, para poder identificar el modelo que mejor se ajusta a la población que se utilizó para el muestreo.

Variables disponibles

Lista las variables de factor y covariable que están seleccionadas en el diálogo principal Regresión lineal Bayesiana. Cuando se añaden o eliminan variables de factor y covariable del diálogo principal, la lista se actualiza en consecuencia.

Variables de prueba

Seleccione las variables de factor/covariable para contrastar desde la lista **Variables disponibles** y añádalas a la lista **Variables de pruebas**.

Nota: La opción **Incluir término de intersección** debe estar seleccionada cuando no hay seleccionado ningún factor o covariable de prueba.

Variables y valores de prueba

Especifique los valores que se van a probar. El número de valores debe cumplir el número de parámetros en el modelo original. Cuando se especifican valores, el primer valor se debe especificar para el término de interceptación (suponer que todos los valores son 0 cuando se han definido explícitamente).

Incluir término de interceptación

Cuando está seleccionado, los términos de interceptación se incluyen en la prueba. De forma predeterminada, el valor no está seleccionado.

Cuando está habilitado, utilice el campo **Valor prueba** para especificar un valor.

Etiqueta de prueba (opcional)

Si lo desea puede especificar una etiqueta para cada prueba. Puede especificar un valor de serie con una longitud máxima de 255 bytes. Sólo se permite una etiqueta por cada prueba F.

ANOVA de un factor Bayesiana

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El procedimiento ANOVA de un factor genera un análisis de varianza de un factor para una variable dependiente cuantitativa respecto a una única variable de factor (la variable independiente). El análisis de la varianza se utiliza para probar la hipótesis de que varias medias son iguales. SPSS Statistics admite factores bayesianos, previas conjugadas y previas no informativas.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > ANOVA de un factor

2. Seleccione una sola variable **Dependiente** numérica en la lista **Variables disponibles**. Debe seleccionar al menos una variable.

3. Seleccione una sola variable de **Factor** para el modelo en la lista **Variables disponibles**. Debe seleccionar como mínimo una variable **Factor**.

4. Seleccione una sola variable no de cadena para que actúe como la **ponderación** de regresión en la lista **Variables disponibles**. El campo de la variable **Ponderación** puede estar vacío.

5. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede

investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.

- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 9. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

15

16

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.

Si lo desea, puede:

- Pulse **Criterios** para especificar el porcentaje de intervalo creíble y valores de método numérico.
- Pulse **A priori** para definir valores de referencia y de distribución a priori de conjugado.
- Pulsar **Factor Bayes** para especificar los valores de factor de Bayes.
- Pulse en **Gráficos** para controlar los gráficos que se generan.

ANOVA unidireccional bayesiana: Criterios

Puede especificar los siguientes criterios de análisis para los modelos ANOVA unidireccional bayesiana.

Porcentaje % de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95%.

Nota: Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayes** o **Utilizar ambos métodos** para el **Análisis Bayesiano**.

Método numérico

Especifique el método numérico utilizado para estimar la integral.

¹⁵ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁶ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Tolerancia

Especifique el valor de tolerancia para los métodos numéricos. El valor predeterminado es 0,000001.

Número máximo de iteraciones

Especifique el número máximo de iteraciones del método. El valor debe ser un entero positivo. El valor predeterminado es 2000.

ANOVA unidireccional bayesiana: Previas

Puede especificar los valores de distribución a priori siguientes para los parámetros de regresión y la varianza de los errores. Las opciones siguientes están disponibles solo cuando la opción **Caracterizar distribución a priori** está seleccionada para el **Análisis Bayesiano**.

Nota: Muchos investigadores pueden cuestionarse la necesidad de especificar un valor previo. Los valores previos de referencia minimizan la preocupación que surge cuando se desborda el valor previo a medida que aumentan los datos. Cuando se especifica información previa informativa, los métodos bayesianos pueden utilizar de forma eficaz la información. El requisito de especificar una prioridad no debe considerarse como un impedimento para utilizar el análisis bayesiano.

Referencia

Cuando está seleccionado, el análisis de referencia genera una inferencia Bayesiana objetiva. Las sentencias inferenciales solo dependen del modelo asumido y los datos disponibles y la distribución a priori que se utiliza para realizar una inferencia es la menos informativa. Ésta es el ajuste predeterminado.

Conjugar

Proporciona opciones para definir distribuciones previas de conjugación. Las previas de conjugación asumen la distribución conjunta Normal, Inversa, Gamma. Aunque las previas de conjugación no son necesarias cuando se realizan actualizaciones Bayesianas, ayudan en los procesos de cálculo.

Previas en varianza de errores

Parámetro de forma

Especifique el parámetro de forma a_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0.

Parámetro de escala

Especifique el parámetro de escala b_0 para la distribución Gamma inversa. Debe especificar un único valor que sea mayor que 0. Cuanto mayor sea el parámetro de escala, más dispersa será la distribución.

Previas en parámetros de regresión

Especifique el factor de media β_0 para las medias de grupo. El número de valores debe cumplir el número de parámetros de regresión, incluido el término de intersección.

La columna **Variables** se llena automáticamente con los niveles del Factor. La columna **Media** no incluye los valores predeterminados.

Pulse **Restablecer** para borrar los valores.

Varianza de matriz de covarianzas: 2×2

Especifique los valores eV_0 en el triángulo inferior de la matriz de varianza-covarianza para la previa normal multivariante. Tenga en cuenta que V_0 debe ser definida semi positiva. Solo se debe especificar el triángulo inferior de la tabla.

Las filas y columnas se llenan automáticamente con los niveles del Factor. Todos los valores de la diagonal son 1; todos los valores fuera de la diagonal son 0.

Pulse **Restablecer** para borrar los valores.

Utilizar matriz de identidad

Cuando se selecciona esta opción, se utiliza la matriz de identidad. No puede especificar los valores V_0 en el triángulo inferior de la matriz de varianza-covarianza para la previa normal multivariante.

ANOVA unidireccional Bayesiana: Factor Bayesiano

Puede especificar el enfoque que se utiliza para estimar el factor Bayesiano para los modelos ANOVA unidireccionales Bayesianos. Las opciones siguientes solo están disponibles cuando está seleccionada la opción de Análisis Bayesiano **Estimar factor Bayesiano** o **Utilizar ambos métodos**.

Cálculo

Especifique el enfoque para estimar factores Bayesianos. El método JZS es el valor predeterminado.

Método JZS

Cuando está seleccionado, invocar el enfoque de Zellner-Siow. Ésta es el ajuste predeterminado.

Método de Zellner

Cuando se selecciona, invoca el enfoque de Zellner y se le pedirá que especifique un único valor anterior $g > 0$ (no hay ningún valor predeterminado).

Método hyper previo

Cuando se selecciona, invoca el enfoque hiper g y se le pedirá que especifique un parámetro de forma α_0 para la distribución gamma inversa. Debe especificar un único valor > 0 (el valor predeterminado es 3).

Método de Rouder

Cuando se selecciona, invoca el enfoque de Rouder y se le pedirá que especifique un parámetro de escalado b_0 para la distribución gamma inversa. Debe especificar un único valor > 0 (el valor predeterminado es 1).

ANOVA unidireccional Bayesiana: Gráficos

Puede controlar los gráficos que se trazan como resultados.

Trazar grupo(s)

Especifique los subgrupos que se van a trazar. Trace la verosimilitud, previa y posterior de las medias de los grupos especificados. La lista **Grupos** es un subconjunto de las categorías de la variable de factor, de modo que el formato debe ser coherente con el tipo de datos y los valores reales del factor.

Varianza de términos de error

Cuando está seleccionado, se traza un gráfico de la varianza de errores. El valor está deseleccionado de forma predeterminada. Esta opción no está disponible cuando **Estimar factor Bayesiano** está seleccionado en el análisis Bayesiano.

Modelos de logaritmo lineal Bayesianos

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

El diseño para probar la independencia de dos factores requiere dos variables categóricas para la construcción de una tabla de contingencia, y realiza una inferencia bayesiana sobre la asociación de fila y columna. Puede estimar los factores bayesianos asumiendo distintos modelos, y caracterizar la distribución posterior deseada simulando el intervalo creíble simultáneo para los términos de interacción.

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > Modelos loglineales

2. Seleccione una única variable de fila no de escala en la lista **Variables disponibles**. Debe seleccionar, al menos, una variable no de escala.
3. Seleccione una única variable de columna no de escala en la lista **Variables disponibles**. Debe seleccionar, al menos, una variable no de escala.
4. Seleccione el **Análisis Bayesiano** deseado:
 - **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.

- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 10. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

17

18

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.

Si lo desea, puede:

- Pulse **Criterios** para especificar el porcentaje de intervalo creíble y valores de método numérico.
- Pulsar **Factor Bayes** para especificar los valores de factor de Bayes.
- Pulsar **Imprimir** para especificar cómo se visualiza el contenido en las tablas de salida.

Modelos de logaritmo lineal Bayesianos: Criterios

Puede especificar los criterios de análisis siguientes para los modelos de logaritmo lineal Bayesianos.

Porcentaje % de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95%.

Método numérico

Especifique el método numérico utilizado para estimar la integral.

Establecer semilla personalizada

Cuando está seleccionado, puede especificar un valor de semilla personalizado en el campo **Semilla**. Especifique un valor de semilla aleatoria. El valor debe ser un entero positivo. De forma predeterminada, se asigna un valor de semilla aleatorio.

Nota: Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayes** o **Utilizar ambos métodos** para el **Análisis Bayesiano**.

¹⁷ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁸ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Tolerancia

Especifique el valor de tolerancia para los métodos numéricos. El valor predeterminado es 0,000001.

Número máximo de iteraciones

Especifique el número máximo de iteraciones del método. El valor debe ser un entero positivo. El valor predeterminado es 2000.

Muestras simuladas para distribución posterior

Especifique el número de muestras que se utilizan para igualar la distribución a priori deseada. El valor predeterminado es 10000.

Formato

Seleccione si las categorías se muestran en orden **Ascendente** o **Descendente**. El orden ascendente es el valor predeterminado.

Modelos de logaritmo lineal bayesiano: factor de Bayes

Puede especificar el modelo presupuesto para los datos observados (Poisson, Multinomial o No parametrizado). La distribución multinomial es el valor de forma predeterminada. Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayesiano** o **Utilizar ambos métodos**.

Modelo Poisson

Cuando está seleccionado, el modelo Poisson se supone para los datos observados.

Modelo multinomial

Cuando está seleccionado, el modelo multinomial se supone para los datos observados. Ésta es el ajuste predeterminado.

Márgenes fijos

Seleccione **Suma final**, **Suma de fila** o **Suma de columna** para especificar los totales de márgenes fijos para la tabla de contingencia. **Suma final** es el valor predeterminado.

Distribución a priori

Especifique el tipo de distribución a priori al estimar el factor Bayesiano.

Conjugar

Seleccione para especificar una distribución a priori de conjugado. Utilice la tabla **Parámetros de forma** para especificar los parámetros de forma a_{rs} para la distribución gamma. Debe especificar los parámetros de forma cuando **Conjugado** está seleccionado como el tipo de distribución a priori.

Cuando se especifica un solo valor, se supone que todos los a_{rs} son iguales a este valor. $rs = 1$ es el valor predeterminado. Si debe especificar más de un valor, puede separar los valores con espacios en blanco.

El número de valores numéricos que se especifican en cada fila y cada columna debe coincidir con la dimensión de la tabla de contingencia. Todos los valores especificados deben ser > 0 .

Pulse **Restablecer** para borrar los valores.

Parámetro de escala

Especifique el parámetro de escala b para la distribución gamma. Debe especificar un solo valor > 0 .

Dirichlet de mezcla

Seleccione para especificar una distribución a priori de Dirichlet de mezcla.

Intrínseco

Seleccione para especificar una distribución a priori intrínseca.

Modelo no parametrizado

Cuando está seleccionado, el modelo no parametrizado se da por supuesto para los datos observados.

Márgenes fijos

Seleccione **Suma de fila** o **Suma de columna** para especificar los totales de márgenes fijos para la tabla de contingencia. **Suma de fila** es el valor predeterminado.

Distribución a priori

Especifique los parámetros para las previas de Dirichlet. Debe especificar los parámetros de **Distribución a priori** cuando está seleccionado el **Modelo no parametrizado**. Cuando se especifica un solo valor, se supone que todos los λ_s son iguales a este valor. $s = 1$ es el valor predeterminado. Si debe especificar más de un valor, puede separar los valores con espacios en blanco. Todos los valores especificados deben ser > 0 . El número de valores numéricos especificados debe coincidir con la dimensión de la fila o columna que no es fija para la tabla de contingencia.

Pulse **Restablecer** para borrar los valores.

Modelos de logaritmo lineal Bayesianos: Imprimir

Puede especificar cómo se muestran los contenidos en las tablas de salida.

Diseño de tabla

Suprimir tabla

Cuando se selecciona esta opción, la tabla de contingencia no se incluye en la salida. El valor no está habilitado de forma predeterminada.

Nota: Los siguientes valores no surten efecto cuando el valor **Suprimir tabla** está habilitado.

Estadísticos

Especifique las estadísticas para las pruebas de independencia.

Chi-cuadrado

Seleccione para calcular la estadística de chi-cuadrado de Pearson, los grados de libertad y la significancia asintótica de 2 lados. Para una tabla de contingencia de 2 por 2, este valor también calcula las estadísticas corregidas de continuidad de Yates y la significación asintótica bilateral. Para una tabla de contingencia de 2 por 2 con, al menos un recuento de celdas previsto < 5 , este valor también calcula la significancia exacta de 1 lado y de 2 lados de la prueba exacta de Fisher.

Razón de verosimilitud

Seleccione esta opción para calcular las estadísticas de la prueba de razón de similitud, grados de libertad y significación sintótica bilateral asociada.

Recuentos

Especifique qué tipos de recuento se incluyen en la tabla de contingencia.

Se ha observado

Seleccione esta opción para incluir recuentos observados en la tabla de contingencia.

Esperado

Seleccione para incluir los recuentos de celdas previstos en la tabla de contingencia.

Porcentajes

Especifique qué tipos de porcentaje están incluidos en la tabla de contingencia.

Fila

Seleccione para incluir los porcentajes de fila en la tabla de contingencia.

Columna

Seleccione esta opción para incluir los porcentajes de columna en la tabla de contingencia.

Total

Seleccione esta opción para incluir porcentajes totales en la tabla de contingencia.

Modelos de ANOVA de medidas repetidas de un factor Bayesiano

Esta característica requiere SPSS Statistics edición Standard o la opción Estadísticas avanzadas.

En los modelos de análisis de varianza (ANOVA) de un factor Bayesiano, se presupone que hay una sola medición por sujeto. Sin embargo, esta suposición no siempre es cierta. No es poco común que un diseño de estudio tenga como objetivo investigar las respuestas de la media en varios puntos de tiempo o condiciones. El procedimiento ANOVA de medidas repetidas unidireccional bayesiano mide un factor del mismo sujeto en cada punto del tiempo o condición distinto, y permite cruzar los sujetos dentro de los niveles. Se da por supuesto que cada sujeto tiene una sola observación para cada punto de tiempo o condición (como tal, la interacción de sujeto-tratamiento no se contabiliza).

1. En los menús seleccione:

Analizar > Estadísticas bayesianas > ANOVA de medidas repetidas de un factor

2. Seleccione como mínimo dos variables **Medidas repetidas** en la lista **Variables disponibles**.

3. Opcionalmente, seleccione una sola variable para que actúe como **Ponderación** de regresión en la lista **Variables disponibles**. El campo de la variable **Ponderación** puede estar vacío.

Nota: La lista de variables disponibles proporciona todas las variables excepto para las variables de cadena.

4. Seleccione el **Análisis Bayesiano** deseado:

- **Caracterizar distribución posterior:** Cuando está seleccionado, la inferencia Bayesiana se realiza desde una perspectiva a la que se ha llegado caracterizando distribuciones posteriores. Puede investigar la distribución posterior marginal de los parámetros de interés integrando los otros parámetros de molestia y, además, construir intervalos creíbles para trazar una inferencia directa. Ésta es el ajuste predeterminado.
- **Estimación del factor Bayes:** cuando se selecciona, la estimación de los factores bayesianos (una de las mejores metodologías en la inferencia bayesiana) constituye un índice natural para comparar las probabilidades marginales entre un nulo y una hipótesis alternativa.

Tabla 11. Umbrales utilizados comúnmente para definir la significación de las pruebas

Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba	Factor Bayesiano	Categoría de prueba
>100	Pruebas extremas para H0	1-3	Pruebas circunstanciales para H0	1/30-1/10	Pruebas sólidas para H1
30-100	Pruebas fehacientes para H0	1	Sin pruebas	1/100-1/30	Pruebas fehacientes para H1
10-30	Pruebas sólidas para H0	1/3-1	Pruebas circunstanciales para H1	1/100	Pruebas extremas para H1
3-10	Pruebas moderadas para H0	1/10-1/3	Pruebas moderadas para H1		

H0: Hipótesis nula

H1: Hipótesis alternativa

19

20

- **Utilizar ambos métodos:** Cuando está seleccionado, se utilizan ambos métodos de inferencia, **Caracterizar distribución posterior** y **Estimar factor Bayesiano**.

¹⁹ Lee, M.D. y Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

²⁰ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Si lo desea, puede:

- Pulse **Criterios** para especificar el porcentaje de intervalo creíble y valores de método numérico.
- Pulsar **Factor Bayes** para especificar los valores de factor de Bayes.
- Pulse **Gráficos** para trazar las distribuciones posteriores de las medias de grupo.

ANOVA de medidas repetidas de un factor Bayesiano: Criterios

Puede especificar los criterios de análisis siguientes para modelos ANOVA de medidas repetidas de un factor Bayesiano.

Porcentaje % de intervalo creíble

Especifique el nivel de significación para calcular intervalos creíbles. El nivel predeterminado es 95%.

Método numérico

Especifique el método numérico utilizado para estimar la integral.

Establecer semilla personalizada

Cuando está seleccionado, puede especificar un valor de semilla personalizada en el campo **Semilla**. El valor predeterminado es 2.000.000. El valor debe ser un entero positivo entre 1 y 2.147.483.647. De forma predeterminada, se asigna un valor de semilla aleatorio.

Número de muestras de Monte Carlo

Especifique el número de puntos que se han muestreado para la aproximación Montecarlo. El valor debe ser un entero positivo entre 10^3 y 10^6 . El valor predeterminado es 30.000.

ANOVA de medidas repetidas de un factor Bayesiano: Factor Bayes

Puede especificar el enfoque que se utiliza para estimar el factor Bayes para los modelos de ANOVA de medidas repetidas de un factor Bayesiano. Las opciones siguientes están disponibles solo cuando está seleccionada la opción **Estimar factor Bayesiano** o **Utilizar ambos métodos**.

Criterios de información bayesiana (BIC)

Utiliza una extensión de la aproximación de BIC al diseño de medidas repetidas para estimar los factores Bayesiano. El valor deriva el tamaño efectivo de la muestra teniendo en cuenta la correlación de medidas repetidas, y sugiere un término de penalización mejorada al estimar el BIC para la selección entre dos modelos que compiten. Ésta es el ajuste predeterminado.

Diseño mixto de Roudier

Utiliza las generalizaciones multivariantes de la distribución de Cauchy como a priori para el tamaño de efecto estandarizado, y un a priori no informativo para la varianza.

Nota: El valor de ponderación de frecuencia global y la ponderación de regresión se ignoran cuando se selecciona esta opción.

ANOVA de medidas repetidas de un factor Bayesiano: Gráficos

Puede controlar los gráficos que se van a generar para ilustrar las distribuciones posteriores de las medias de grupo. La tabla enumera todas las variables que se seleccionan como medidas repetidas en el diálogo Variables. Seleccione las variables de medidas repetidas que se van a representar en el gráfico.

Regresión Ridge del kernel

La regresión Ridge del kernel es un procedimiento de extensión que utiliza la clase Python `sklearn.kernel_ridge.KernelRidge` para estimar los modelos de regresión ridge del kernel. Los modelos de regresión ridge del kernel son modelos de regresión no paramétrica que son capaces de modelar relaciones lineales y no lineales entre las variables predictoras y los resultados. Los resultados pueden ser muy sensibles a las opciones de los hiperparámetros del modelo. La regresión ridge del kernel facilita la elección de los valores de hiperparámetro a través de la validación cruzada de k en las cuadrículas especificadas de valores utilizando la clase `sklearn.model_selection.GridSearchCV`.

Ejemplo

Estadísticos

Additive_CHI2, CHI2, Coseno, Laplaciana, Lineal, Polinómica, RBF, Sigmoide, Alfa, Gamma, Coef0, Grado, validación cruzada, observados respecto a pronosticados, residuos respecto a pronosticados, coeficientes de ponderación dual, coeficientes de ponderación del espacio del kernel.

Consideraciones sobre los datos

Datos

- Puede especificar cualquiera de o todas las ocho funciones de kernel diferentes.
- La función de kernel seleccionada determina qué hiperparámetros están activos.
- Los hiperparámetros incluyen alfa para la regularización ridge que son comunes a todos los kernels además de otros tres hiperparámetros para cada función de kernel específica.
- Cuando se especifican varios submandatos de kernel, o se especifica más de un valor para cualquier parámetro, se realiza una búsqueda de cuadrícula con validación cruzada para evaluar modelos y se selecciona el mejor modelo de ajuste basado en los datos retenidos.
- La extensión acepta variables de división del procedimiento Segmentar archivo y ponderaciones utilizando el procedimiento Ponderación de casos.
- Cuando se incluyen ponderaciones, se utilizan en la creación de valores ajustados en todos los análisis. Debido a limitaciones en el método de puntuación en la clase **sklearn.model_selection.GridSearchCV**, las evaluaciones de validación cruzada que se utilizan para la selección de modelos no se ponderan.

Supuestos

Obtención de una regresión ridge de kernel

1. En los menús seleccione:

Analizar > Regresión > Kernel Ridge...

2. Seleccione una variable **Dependiente**.

3. Seleccione una o más variables **Independientes**.

4. El valor predeterminado de **Modelo único** se utiliza cuando sólo se especifica un valor para cada parámetro de función de kernel. Cuando se selecciona el valor **Modelo único**, no puede especificar funciones y ponderaciones adicionales de **Kernel(s)** que se aplican por completo a lo largo del análisis, la evaluación y la puntuación de los resultados. También puede utilizar los controles de flecha arriba y abajo para reorganizar las funciones del kernel.

Opcionalmente, seleccione **Selección de modelos** en la lista **Modo**.

Cuando se selecciona **Selección de modelos** en la lista **Modo**, puede añadir varias funciones de kernel a la lista de **Kernel(s)**.

a. Pulse el control de adición (+) para incluir funciones adicionales del kernel.

b. Pulse la celda vacía en la columna **Kernel** para seleccionar una función de kernel.

c. Efectúe una doble pulsación en cualquier celda de fila de función de kernel para especificar valores de parámetros de función de kernel para la columna correspondiente (**Alfa, Gamma, Coef0, Grado**). Para obtener más información, consulte “Parámetros de kernel” en la [página 123](#). Los parámetros de ajuste de la función de kernel predeterminados se listan a continuación.

Additive_CHI2

ALFA=1 GAMMA=1

CHI2

ALFA=1 GAMMA=1

Coseno

ALFA = 1

Laplaciana

ALFA= 1 GAMMA=1/p

Lineal

La función de kernel predeterminada. ALFA = 1

Polinómica

ALFA= 1 GAMMA=1/p COEF0 = 1 GRADO= 3

RBF

ALFA= 1 GAMMA=1/p

Sigmoide

ALFA= 1 GAMMA=1/p COEF0 = 1

Nota: Cuando se especifica más de un valor para cualquier parámetro de función de kernel, se realiza una búsqueda de cuadrícula con validación cruzada para evaluar modelos, y se selecciona el mejor modelo de ajuste basado en los datos retenidos.

- De forma opcional, pulse **Opciones** para especificar el número de pliegues de validación cruzada, opciones de visualización, valores de trazado y elementos para guardar. Para obtener más información, consulte [“Regresión Ridge de kernel: Opciones”](#) en la página 123.
- Pulse **Aceptar**.

Parámetros de kernel

El diálogo **Parámetros del kernel** proporciona opciones para especificar valores de parámetros de función de kernel individuales y para especificar que la selección de modelo se realiza utilizando una búsqueda de cuadrícula sobre las combinaciones de kernels y los valores de parámetros de cuadrícula especificados.

Especificar parámetros individuales

Habilite el valor para especificar valores para el parámetro de función de kernel seleccionado.

- Especifique un valor y pulse **Añadir** para incluir el valor en el parámetro de función del kernel.
- Seleccione un valor de parámetro y pulse **Cambiar** para actualizar el valor.
- Seleccione un valor de parámetro y pulse **Eliminar** para suprimir el valor.

Especificar parámetros de cuadrícula

Habilite el valor para especificar que la selección de modelo se realiza utilizando una búsqueda de cuadrícula sobre las combinaciones de kernels y los valores de parámetros de cuadrícula especificados.

Regresión Ridge de kernel: Opciones

El diálogo **Gráficos** proporciona opciones para especificar el número de pliegues de validación cruzada, opciones de visualización, valores de gráficos y elementos para guardar.

Número de pliegues de validación cruzada

El número de divisiones o pliegues en la validación cruzada con la búsqueda de cuadrícula para la selección de modelos. Especifique un valor entero mayor que 1. El valor predeterminado es 5. El valor sólo está disponible cuando se elige **Selección de modelos** como **Modo** en el diálogo principal **Regresión Ridge del kernel**.

Visualizar

Proporciona opciones para especificar qué salida se debe visualizar cuando está en vigor la validación cruzada.

Mejor

El valor predeterminado sólo muestra los resultados básicos para el mejor modelo elegido.

Comparar

Muestra los resultados básicos para todos los modelos evaluados.

Comparar modelos y pliegues

Muestra los resultados completos de cada división o pliegue para cada modelo evaluado.

Gráfico

Proporciona opciones para especificar gráficos de valores observados o residuales frente a valores pronosticados.

Observados frente a pronosticados

Muestra un diagrama de dispersión de los valores observados frente a los previstos para el modelo especificado o mejor.

Residuales frente a predichos

Muestra un diagrama de dispersión de los residuos frente a los valores pronosticados para el modelo especificado o mejor.

Guardar

La tabla proporciona opciones para especificar variables para guardar en el conjunto de datos activo.

Valores pronosticados

Guarda los valores pronosticados del modelo especificado o mejor en el conjunto de datos activo. Se puede incluir un nombre de variable opcional.

Residuos

Guarda los residuos de las predicciones del modelo especificado o mejor en el conjunto de datos activo. Se puede incluir un nombre de variable opcional.

Coefficientes duales

Guarda los coeficientes de ponderación de espacio dual o de kernel del modelo especificado en el conjunto de datos activo. Se puede incluir un nombre de variable opcional. El valor no está disponible cuando se elige **Selección de modelos** como **Modo** en el diálogo principal **Regresión Ridge del kernel**.

Avisos

Esta información se ha desarrollado para productos y servicios que se ofrecen en EE.UU. Este material de IBM podría estar disponible en otros idiomas. Sin embargo, puede que se le exija poseer una copia del producto o de una versión del producto en ese idioma para poder acceder a él.

Es posible que IBM no ofrezca los productos, servicios o características descritos en este documento en otros países. Póngase en contacto con el representante local de IBM, que le informará sobre los productos y servicios disponibles actualmente en su área. Cualquier referencia a un producto, programa o servicio de IBM no pretende afirmar ni implicar que solo se puede utilizar dicho producto, programa o servicio de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no infrinja ninguno de los derechos intelectuales de IBM. Sin embargo, es responsabilidad del usuario evaluar y verificar el funcionamiento de cualquier producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patentes pendientes que cubran el tema principal descrito en este documento. El suministro de este documento no le otorga ninguna licencia sobre dichas patentes. Puede enviar consultas sobre licencias, por escrito, a:

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119Armonk, NY 10504-1785Estados Unidos

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Licencia de Propiedad intelectual

Legal and Intellectual Property Law

IBM Japan Ltd.

19-21, Nihonbashi-Hakozakicho, Chuo-kuTokio 103-8510, Japón

INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGUNA CLASE, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO, PERO SIN LIMITARSE A, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O ADECUACIÓN A UN FIN DETERMINADO. Algunas jurisdicciones no permiten la renuncia de garantías expresas o implícitas en determinadas transacciones, por lo que esta declaración podría no ser aplicable a su caso.

Esta publicación puede contener inexactitudes técnicas o errores tipográficos. Periódicamente se efectúan cambios en la información aquí contenida; estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar en cualquier momento mejoras o cambios en los productos o programas descritos en esta publicación sin previo aviso.

Las referencias en esta información a sitios web que no son de IBM se proporcionan solo por comodidad y de ningún modo suponen un aval de dichos sitios. La información de esos sitios web no forma parte de la información de este producto de IBM y la utilización de esos sitios web se realiza bajo la responsabilidad del usuario.

IBM puede utilizar o distribuir cualquier información que proporcione de la forma que crea adecuada sin incurrir por ello en ninguna obligación con el cliente.

Los licenciarios de este programa que deseen tener información sobre él con el fin de permitir: (i) el intercambio de información entre programas creados independientemente y otros programas (incluido el presente programa) y (ii) el uso recíproco de la información que se ha intercambiado, deben ponerse en contacto con:

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119Armonk, NY 10504-1785Estados Unidos

Dicha información puede estar disponible, sujeta a los términos y condiciones correspondientes, incluidos en algunos casos, el pago de una tarifa.

El programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible lo proporciona IBM bajo los términos de las Condiciones generales de IBM, Acuerdo internacional de programas bajo licencia de IBM o cualquier otro acuerdo equivalente entre ambas partes.

Los ejemplos de cliente y datos de rendimiento citados se presentan solo con fines ilustrativos. Los resultados de rendimiento reales pueden variar en función de las configuraciones específicas y de las condiciones de funcionamiento.

La información relativa a los productos que no son de IBM se ha obtenido de los proveedores de dichos productos, de sus anuncios publicados o de otras fuentes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión del rendimiento, la compatibilidad ni ninguna otra declaración relacionada con productos no IBM. Las preguntas sobre las prestaciones de productos que no son de IBM deben dirigirse a los proveedores de dichos productos.

Las declaraciones relativas a la dirección o intenciones futuras de IBM pueden cambiar o ser retiradas sin previo aviso, y solo representan propósitos y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos de la forma más completa posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Todos estos nombres son ficticios y cualquier parecido con los nombres de personas o empresas reales es pura coincidencia.

LICENCIA DE DERECHOS DE AUTOR:

Esta información contiene programas de aplicación de ejemplo en lenguaje fuente, que se utilizan para ilustrar las técnicas de programación y distintas plataformas operativas. Puede copiar, modificar y distribuir estos programas de muestra de cualquier modo sin pagar a IBM con el fin de desarrollar, utilizar, comercializar o distribuir programas de aplicación que se ajusten a la interfaz de programación de aplicaciones para la plataforma operativa para la que se ha escrito el código de muestra. Estos ejemplos no se han probado exhaustivamente en todas las condiciones. Por lo tanto, IBM no puede garantizar ni dar por supuesta la fiabilidad, la capacidad de servicio ni la funcionalidad de estos programas. Los programas de muestra se proporcionan "TAL CUAL" sin garantía de ningún tipo. IBM no se hará responsable de los daños derivados del uso de los programas de muestra.

Cada copia o cualquier parte de estos programas de muestra o cualquier trabajo derivado debe incluir un aviso de copyright, de esta manera:

© Copyright IBM Corp. 2021. Las partes de este código se derivan de IBM Corp. Programas de ejemplo.

© Copyright IBM Corp. 1989 - 2021. Reservados todos los derechos.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas registradas o marcas comerciales registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras empresas. Hay disponible una lista actualizada de las marcas registradas de IBM en la web, en "Copyright and trademark information", en www.ibm.com/legal/copytrade.shtml.

Adobe, el logotipo de Adobe, PostScript y el logotipo de PostScript son marcas registradas de Adobe Systems Incorporated en los Estados Unidos o en otros países.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas registradas o marcas comerciales de Intel Corporation o de sus subsidiarias en los Estados Unidos y en otros países.

Linux es una marca registrada de Linus Torvalds en los Estados Unidos y/o en otros países.

Microsoft, Windows, Windows NT y el logotipo de Windows son marcas registradas de Microsoft Corporation en Estados Unidos o en otros países.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas registradas y logotipos basados en Java son marcas comerciales o marcas registradas de Oracle y/o sus filiales.

Índice

A

- análisis de covarianza
 - en GLM multivariante [1](#)
- análisis de la varianza
 - en modelos mixtos lineales generalizados [52](#)
- análisis de supervivencia
 - en Kaplan-Meier [76](#)
 - en la regresión de Cox [79](#)
 - en las tablas de mortalidad [74](#)
 - en Regresión Ridge de kernel [121](#)
 - Regresión de Cox dependiente del tiempo [82](#)
- análisis de varianza
 - en los componentes de la varianza [20](#)
- análisis loglineal
 - Análisis loglineal general [68](#)
 - Análisis loglineal logit [71](#)
 - en modelos mixtos lineales generalizados [52](#)
- Análisis loglineal general
 - almacenamiento de valores pronosticados [70](#)
 - características adicionales del mandato [71](#)
 - contrastes [68](#)
 - covariables de casilla [68](#)
 - criterios [70](#)
 - distribución de recuentos de casillas [68](#)
 - especificación de modelo [69](#)
 - estructuras de casilla [68](#)
 - factores [68](#)
 - gráficos [70](#)
 - guardar variables [70](#)
 - intervalos de confianza [70](#)
 - opciones de visualización [70](#)
 - residuos [70](#)
- Análisis loglineal logit
 - almacenamiento de variables [73](#)
 - contrastes [71](#)
 - covariables de casilla [71](#)
 - criterios [73](#)
 - distribución de recuentos de casillas [71](#)
 - especificación de modelo [72](#)
 - estructuras de casilla [71](#)
 - factores [71](#)
 - gráficos [73](#)
 - intervalos de confianza [73](#)
 - opciones de visualización [73](#)
 - residuos [73](#)
 - valores pronosticados [73](#)
- Análisis loglineal: Selección de modelo
 - características adicionales del mandato [68](#)
 - definición de los rangos del factor [67](#)
 - Modelos de [67](#)
 - opciones [68](#)
- análisis probit
 - modelos lineales mixtos generalizados [52](#)
- ANOVA
 - en GLM multivariante [1](#)
 - en MLG medidas repetidas [9](#)

ANOVA multivariada [1](#)

B

- bondad de ajuste
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
- Bonferroni
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)

C

- C de Dunnett
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
- casos censurados
 - en Kaplan-Meier [76](#)
 - en la regresión de Cox [79](#)
 - en las tablas de mortalidad [74](#)
- categoría de referencia
 - en ecuaciones de estimación generalizadas [45](#)
 - en modelos lineales generalizados [33](#)
- clase generadora
 - en el análisis loglineal de selección de modelo [67](#)
- Componentes de la varianza
 - almacenamiento de resultados [21](#)
 - características adicionales del mandato [22](#)
 - modelo [20](#)
 - opciones [20](#)
- construir términos [3](#), [12](#), [20](#), [68](#), [70](#), [73](#)
- contraste de multiplicador de Lagrange
 - en modelos lineales generalizados [36](#)
- contrastes
 - en el análisis loglineal general [68](#)
 - en el análisis loglineal logit [71](#)
 - en la regresión de Cox [80](#)
- convergencia de los parámetros
 - en ecuaciones de estimación generalizadas [46](#)
 - en modelos lineales generalizados [34](#)
 - en modelos lineales mixtos [27](#)
- convergencia del log-verosimilitud
 - en ecuaciones de estimación generalizadas [46](#)
 - en modelos lineales generalizados [34](#)
- convergencia del logaritmo de la verosimilitud
 - en modelos lineales mixtos [27](#)
- Convergencia hessiana
 - en ecuaciones de estimación generalizadas [46](#)
 - en modelos lineales generalizados [34](#)
- covariables
 - en la regresión de Cox [80](#)
- covariables de cadena
 - en la regresión de Cox [80](#)
- covariables segmentadas dependientes del tiempo
 - en la regresión de Cox [82](#)

D

- descomposición jerárquica
 - en los componentes de la varianza [21](#)
- diferencia honestamente significativa de Tukey
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
- diferencia menos significativa
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
- Distancia de Cook
 - en MLG [7](#)
 - en MLG Medidas repetidas [16](#)
 - en modelos lineales generalizados [38](#)
- distribución binomial
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución binomial negativa
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución De Gauss inversa
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución de Poisson
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución gamma
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución multinomial
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución normal
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- distribución Tweedie
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- DMS de Fisher
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)

E

- Ecuaciones de estimación generalizadas
 - categoría de referencia para respuesta binaria [45](#)
 - criterios de estimación [46](#)
 - especificación de modelo [46](#)
 - estadísticos [48](#)
 - exportación del modelo [51](#)
 - guardar variables en el conjunto de datos activo [50](#)
 - medias marginales estimadas [49](#)
 - opciones para factores categóricos [45](#)
 - predictores [45](#)
 - respuesta [44](#)
 - tipo de modelo [42](#)
 - valores iniciales [47](#)
- efectos aleatorios
 - en modelos lineales mixtos [26](#)
- efectos fijos
 - en modelos lineales mixtos [25](#)
- eliminación hacia atrás
 - en el análisis loglineal de selección de modelo [66](#)
- EMNCI

EMNCI (continuación)

- en los componentes de la varianza [20](#)
- error estándar
 - en MLG [7](#)
 - en MLG Medidas repetidas [16](#)
- estadísticas descriptivas
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
 - en modelos lineales mixtos [28](#)
- estadístico de Wald
 - en el análisis loglineal general [68](#)
 - en el análisis loglineal logit [71](#)
- estimación de la máxima verosimilitud
 - en los componentes de la varianza [20](#)
- estimación de la máxima verosimilitud restringida
 - en los componentes de la varianza [20](#)
- estimaciones de los parámetros
 - en el análisis loglineal de selección de modelo [68](#)
 - en el análisis loglineal general [68](#)
 - en modelos lineales mixtos [28](#)
- estimaciones de parámetros
 - en ecuaciones de estimación generalizadas [48](#)
 - en el análisis loglineal logit [71](#)
 - en modelos lineales generalizados [36](#)
- estructuras de covarianza
 - en modelos lineales mixtos [87](#)

F

- F múltiple de Ryan-Einot-Gabriel-Welsch
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
- factores
 - en MLG medidas repetidas [12](#)
- frecuencias
 - en el análisis loglineal de selección de modelo [68](#)
- función de enlace
 - modelos mixtos lineales generalizados [54](#)
- función de enlace binomial negativa
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace Cauchit acumulada
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace complementaria log
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace de identidad
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace de poder de probabilidad
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace de potencia
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace log
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace log-log complementaria
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
- función de enlace log-log complementaria acumulada
 - en ecuaciones de estimación generalizadas [42](#)

- función de enlace log-log complementaria acumulada (*continuación*)
 - en modelos lineales generalizados [29](#)
 - función de enlace log-log negativa
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
 - función de enlace log-log negativa acumulada
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
 - función de enlace logit
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
 - función de enlace logit acumulada
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
 - función de enlace probit
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
 - función de enlace probit acumulada
 - en ecuaciones de estimación generalizadas [42](#)
 - en modelos lineales generalizados [29](#)
 - función de supervivencia
 - en las tablas de mortalidad [74](#)
 - función estimable general
 - en ecuaciones de estimación generalizadas [48](#)
 - funciones estimables generales
 - en modelos lineales generalizados [36](#)
- G**
- GLM multivariante [1](#)
 - GLM Multivariante
 - covariables [1](#)
 - factores [1](#)
 - gráficos de perfil [5](#)
 - pruebas post hoc [5](#)
 - variable dependiente [1](#)
 - GLOR
 - en el análisis loglineal general [68](#)
 - gráficos
 - en el análisis loglineal general [70](#)
 - en el análisis loglineal logit [73](#)
 - gráficos de perfil
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [14](#)
 - gráficos de probabilidad normal
 - en el análisis loglineal de selección de modelo [68](#)
 - GT2 de Hochberg
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
- H**
- historial de iteraciones
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
 - en modelos lineales mixtos [27](#)
- I**
- información de los niveles del factor
 - en modelos lineales mixtos [28](#)
 - información del modelo
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
 - en el análisis loglineal general [70](#)
 - en el análisis loglineal logit [73](#)
 - en modelos lineales mixtos [28](#)
 - iteraciones
 - en ecuaciones de estimación generalizadas [46](#)
 - en el análisis loglineal de selección de modelo [68](#)
 - en modelos lineales generalizados [34](#)
- J**
- intervalos de confianza
 - en el análisis loglineal general [70](#)
 - en el análisis loglineal logit [73](#)
 - en modelos lineales mixtos [28](#)
- K**
- Kaplan-Meier
 - almacenamiento de nuevas variables [78](#)
 - características adicionales del mandato [78](#)
 - comparación de niveles del factor [77](#)
 - cuartiles [78](#)
 - definición de eventos [77](#)
 - ejemplo [76](#)
 - estadísticos [76, 78](#)
 - gráficos [78](#)
 - media y mediana de tiempos de supervivencia [78](#)
 - tablas de supervivencia [78](#)
 - tendencia lineal para los niveles del factor [77](#)
 - variables de estado de supervivencia [77](#)
- L**
- log-razón de las ventajas generalizada
 - en el análisis loglineal general [68](#)
- M**
- matriz de correlaciones
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
 - en modelos lineales mixtos [28](#)
 - matriz de covarianzas
 - en ecuaciones de estimación generalizadas [46, 48](#)
 - en MLG [7](#)
 - en modelos lineales generalizados [34, 36](#)
 - en modelos lineales mixtos [28](#)
 - matriz de covarianzas de efectos aleatorios
 - en modelos lineales mixtos [28](#)
 - matriz de covarianzas de los parámetros
 - en modelos lineales mixtos [28](#)
 - matriz de covarianzas residuales
 - en modelos lineales mixtos [28](#)
 - matriz de los coeficientes de los contrastes
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
 - matriz L
 - en ecuaciones de estimación generalizadas [48](#)
 - en modelos lineales generalizados [36](#)
 - medias marginales estimadas
 - en ecuaciones de estimación generalizadas [49](#)
 - en modelos lineales generalizados [37](#)
 - en modelos lineales mixtos [28](#)
 - método de Newton-Raphson
 - en el análisis loglineal general [68](#)
 - en el análisis loglineal logit [71](#)
 - MLG

- MLG (*continuación*)
 - almacenamiento de matrices [7](#)
 - guardar variables [7](#)
 - MLG Medidas repetidas
 - características adicionales del mandato [18](#)
 - definir factores [12](#)
 - gráficos de perfil [14](#)
 - guardar variables [16](#)
 - modelo [12](#)
 - pruebas post hoc [15](#)
 - modelo de riesgos proporcionales
 - en la regresión de Cox [79](#)
 - modelo lineal general
 - modelos lineales mixtos generalizados [52](#)
 - modelo lineal generalizado
 - en modelos mixtos lineales generalizados [52](#)
 - modelos factoriales completos
 - en los componentes de la varianza [20](#)
 - en MLG medidas repetidas [12](#)
 - modelos jerárquicos
 - modelos lineales mixtos generalizados [52](#)
 - modelos lineales generalizados
 - medias marginales estimadas [60](#)
 - Modelos lineales generalizados
 - categoría de referencia para respuesta binaria [33](#)
 - criterios de estimación [34](#)
 - distribución [29](#)
 - especificación de modelo [34](#)
 - estadísticos [36](#)
 - exportación del modelo [39](#)
 - función de enlace [29](#)
 - guardar variables en el conjunto de datos activo [38](#)
 - medias marginales estimadas [37](#)
 - opciones para factores categóricos [33](#)
 - predictores [33](#)
 - respuesta [32](#)
 - tipos de modelos [29](#)
 - valores iniciales [35](#)
 - Modelos lineales mixtos
 - características adicionales del mandato [29](#)
 - construir términos [25](#)
 - crear términos [25](#)
 - criterios de estimación [27](#)
 - efectos aleatorios [26](#)
 - efectos fijos [25](#)
 - estructura de covarianza [87](#)
 - guardar variables [29](#)
 - medias marginales estimadas [28](#)
 - modelo [28](#)
 - términos de interacción [25](#)
 - modelos lineales mixtos generalizados
 - bloque de efectos aleatorios [58](#)
 - covarianzas de los efectos aleatorios [64](#)
 - efectos fijos [63](#)
 - offset [59](#)
 - ponderación de análisis [59](#)
 - predicho por observado [63](#)
 - resumen de modelos [62](#)
 - modelos logit multinomiales [71](#)
 - modelos loglineales jerárquicos [66](#)
 - modelos longitudinales
 - modelos lineales mixtos generalizados [52](#)
 - modelos mixtos
 - lineal [22](#)
 - modelos mixtos (*continuación*)
 - modelos lineales mixtos generalizados [52](#)
 - modelos mixtos lineales generalizados
 - coeficientes fijos [64](#)
 - distribución de destino [54](#)
 - efectos aleatorios [58](#)
 - efectos fijos [56](#)
 - estructura de datos [63](#)
 - exportación del modelo [61](#)
 - función de enlace [54](#)
 - guardar campos [61](#)
 - medias estimadas [65](#)
 - parámetros de covarianza [65](#)
 - tabla de clasificación [63](#)
 - términos personalizados [57](#)
 - vista de modelo [62](#)
 - modelos multinivel
 - modelos lineales mixtos generalizados [52](#)
 - modelos personalizados
 - en el análisis loglineal de selección de modelo [67](#)
 - en los componentes de la varianza [20](#)
 - en MLG medidas repetidas [12](#)
 - modelos saturados
 - en el análisis loglineal de selección de modelo [67](#)
- N**
- Newman-Keuls
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
- P**
- parámetro de escala
 - en ecuaciones de estimación generalizadas [46](#)
 - en modelos lineales generalizados [34](#)
 - previas de los efectos aleatorios
 - en los componentes de la varianza [20](#)
 - prueba b de Tukey
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
 - prueba de Breslow
 - en Kaplan-Meier [77](#)
 - Prueba de comparación por parejas de Gabriel
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
 - Prueba de comparación por parejas de Games y Howell
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
 - prueba de Gehan
 - en las tablas de mortalidad [76](#)
 - prueba de log rango
 - en Kaplan-Meier [77](#)
 - prueba de parámetros de covarianza
 - en modelos lineales mixtos [28](#)
 - prueba de rangos múltiples de Duncan
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
 - prueba de Scheffé
 - en GLM multivariante [5](#)
 - en MLG medidas repetidas [15](#)
 - prueba de Tarone-Ware
 - en Kaplan-Meier [77](#)

prueba de Wilcoxon
 en las tablas de mortalidad [76](#)
prueba t de Dunnett
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
prueba t de Sidak
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
prueba t de Waller-Duncan
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
puntuación
 en modelos lineales mixtos [27](#)
Puntuación de Fisher
 en modelos lineales mixtos [27](#)

R

R-E-G-W F
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
R-E-G-W Q
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
rango múltiple de Ryan-Einot-Gabriel-Welsch
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
razón de las ventajas
 en el análisis loglineal general [68](#)
referencias cruzadas en tablas
 en el análisis loglineal de selección de modelo [66](#)
Regresión de Cox
 características adicionales del mandato [82](#)
 contrastes [80](#)
 covariables [79](#)
 covariables categóricas [80](#)
 covariables de cadena [80](#)
 covariables dependientes del tiempo [82](#), [83](#)
 definición de eventos [82](#)
 DfBetas [81](#)
 ejemplo [79](#)
 entrada o eliminación por pasos [81](#)
 estadísticos [79](#), [81](#)
 función de riesgo [81](#)
 función de supervivencia [81](#)
 funciones de línea base [81](#)
 gráficos [80](#)
 guardar variables nuevas [81](#)
 iteraciones [81](#)
 residuos parciales [81](#)
 variable del estado de supervivencia [82](#)
Regresión de Poisson
 en el análisis loglineal general [68](#)
 modelos lineales mixtos generalizados [52](#)
regresión logística
 modelos lineales mixtos generalizados [52](#)
regresión logística multinomial
 modelos lineales mixtos generalizados [52](#)
regresión multivariada [1](#)
Regresión Ridge del kernel
 gráficos [123](#)
 guardar [123](#)
 parámetros [123](#)
 parámetros de cuadrícula [123](#)

Regresión Ridge del kernel (*continuación*)
 pliegues de validación cruzada [123](#)
 visualización [123](#)
residuos
 en ecuaciones de estimación generalizadas [50](#)
 en el análisis loglineal de selección de modelo [68](#)
 en el análisis loglineal general [70](#)
 en el análisis loglineal logit [73](#)
 en modelos lineales generalizados [38](#)
 en modelos lineales mixtos [29](#)
residuos de desviación
 en modelos lineales generalizados [38](#)
residuos de Pearson
 en ecuaciones de estimación generalizadas [50](#)
 en modelos lineales generalizados [38](#)
residuos de verosimilitud
 en modelos lineales generalizados [38](#)
residuos eliminados
 en MLG [7](#)
 en MLG Medidas repetidas [16](#)
residuos estandarizados
 en MLG [7](#)
 en MLG Medidas repetidas [16](#)
residuos no estandarizados
 en MLG [7](#)
 en MLG Medidas repetidas [16](#)
resumen de procesamiento de casos
 en ecuaciones de estimación generalizadas [48](#)
 en modelos lineales generalizados [36](#)
Ridge de kernel
 alpha [121](#)
 coef0 [121](#)
 gamma [121](#)
 grado [121](#)
 modelo único [121](#)
 selección de modelos [121](#)

S

separación
 en ecuaciones de estimación generalizadas [46](#)
 en modelos lineales generalizados [34](#)
Student-Newman-Keuls
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
subdivisión por pasos
 en ecuaciones de estimación generalizadas [46](#)
 en modelos lineales generalizados [34](#)
 en modelos lineales mixtos [27](#)
suma de cuadrados
 en los componentes de la varianza [21](#)
 en modelos lineales mixtos [25](#)

T

T2 de Tamhane
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
T3 de Dunnett
 en GLM multivariante [5](#)
 en MLG medidas repetidas [15](#)
tablas de contingencia
 en el análisis loglineal general [68](#)

- Tablas de mortalidad
 - características adicionales del mandato [76](#)
 - comparación de niveles del factor [76](#)
 - ejemplo [74](#)
 - estadísticos [74](#)
 - función de supervivencia [74](#)
 - gráficos [76](#)
 - prueba de Wilcoxon (Gehan) [76](#)
 - supresión de la presentación de tablas [76](#)
 - tasa de riesgo [74](#)
 - variables de estado de supervivencia [75](#)
 - variables del factor [76](#)
- tasa de riesgo
 - en las tablas de mortalidad [74](#)
- términos anidados
 - en ecuaciones de estimación generalizadas [46](#)
 - en modelos lineales generalizados [34](#)
 - en modelos lineales mixtos [25](#)
- términos de interacción
 - en modelos lineales mixtos [25](#)
- tolerancia para la singularidad
 - en modelos lineales mixtos [27](#)

V

- valores de influencia
 - en MLG [7](#)
 - en MLG Medidas repetidas [16](#)
 - en modelos lineales generalizados [38](#)
- valores pronosticados
 - en el análisis loglineal general [70](#)
 - en el análisis loglineal logit [73](#)
 - en modelos lineales mixtos [29](#)
- valores pronosticados fijos
 - en modelos lineales mixtos [29](#)
- valores pronosticados ponderados
 - en MLG [7](#)
 - en MLG Medidas repetidas [16](#)
- variables de medidas repetidas
 - en modelos lineales mixtos [23](#)
- variables de sujetos
 - en modelos lineales mixtos [23](#)
- vista de modelo
 - en modelos mixtos lineales generalizados [62](#)

