

IBM SPSS Regression 29

IBM

Nota

Antes de utilizar esta información y el producto al que se refiere, lea la información en [“Avisos” en la página 41](#).

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. Regresión.....	1
Selección de un procedimiento para la regresión logística binaria.....	1
Regresión lineal Ridge.....	2
Regresión de cresta lineal: Opciones.....	2
Regresión logística.....	5
Regresión logística: Establecer regla.....	6
Métodos de selección de variables en el análisis de regresión logística.....	6
Regresión logística: Definir variables categóricas.....	7
Regresión logística: Guardar nuevas variables.....	7
Regresión logística: Opciones.....	8
Características adicionales del comando LOGISTIC REGRESSION.....	9
Regresión logística multinomial.....	9
Regresión logística multinomial.....	10
Categoría de referencia de regresión logística multinomial.....	11
Regresión logística multinomial: Estadísticos.....	11
Regresión logística multinomial: Criterios.....	12
Opciones de Regresión logística multinomial.....	12
Regresión logística multinomial: Guardar.....	13
Características adicionales del comando NOMREG.....	13
Regresión probit.....	13
Análisis probit: Definir rango.....	14
Análisis probit: Opciones.....	14
Características adicionales del comando PROBIT.....	15
Regresión cuantil.....	15
Regresión cuantil: Criterios.....	17
Regresión cuantil: Modelo.....	18
Regresión cuantil: Visualización.....	19
Regresión cuantil: Guardar.....	20
Regresión cuantil: Exportar.....	20
Regresión no lineal.....	21
Lógica condicional (Regresión no lineal).....	22
Regresión no lineal: Parámetros.....	23
Modelos comunes de regresión no lineal.....	23
Regresión no lineal: Función de pérdida.....	24
Regresión no lineal: Restricciones para los parámetros.....	24
Regresión no lineal: Guardar variables nuevas.....	24
Regresión no lineal: Opciones.....	25
Regresión no lineal: Interpretación de los resultados.....	25
Características adicionales del comando NLR.....	25
Estimación ponderada.....	26
Estimación ponderada: Opciones.....	27
Características adicionales del comando WLS.....	27
Regresión de cuadrados mínimos de dos etapas.....	27
Regresión por mínimos cuadrados en dos fases: Opciones.....	29
Características adicionales del comando 2SLS.....	29
Esquemas de codificación de variables categóricas.....	29
Desviación.....	29
Sencillo.....	29
Helmert.....	30
Diferencia.....	30
Polinómico.....	30

Repetido.....	31
Especial.....	31
Indicador.....	32
Regresión Ridge del kernel.....	32
Parámetros de kernel.....	34
Regresión Ridge de kernel: Opciones.....	34
Modelos de tiempo de anomalía acelerada paramétrica.....	35
Modelos de tiempo de anomalía acelerada paramétrica: Criterios.....	36
Modelos de tiempo de anomalía acelerada paramétrica: Modelo.....	36
Modelos de tiempo de anomalía acelerada paramétrica: Estimación.....	37
Modelos de tiempo de anomalía acelerada paramétrica: Imprimir.....	38
Modelos de tiempo de anomalía acelerada paramétrica: predecir.....	39
Modelos de tiempo de anomalía acelerada paramétrica: Gráfico.....	40
Modelos de tiempo de anomalía acelerada paramétrica: Exportar.....	40
Avisos.....	41
Marcas comerciales.....	42
Índice.....	45

Capítulo 1. Regresión

Se han incluido las características de regresión siguientes en SPSS Statistics Standard Edition o la opción Regresión.

Selección de un procedimiento para la regresión logística binaria

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

Regresión logística

Proporciona las características exclusivas siguientes:

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

Regresión logística multinomial

Proporciona las características exclusivas siguientes:

- Pruebas chi-cuadrado de Pearson y de desviación sobre la bondad de ajuste del modelo
- Especificación de subpoblaciones para el agrupamiento de los datos, para las pruebas de bondad de ajuste
- Listado de las frecuencias, frecuencias pronosticadas y residuos por subpoblaciones
- Corrección de las estimaciones de la varianza por sobredispersión
- Matriz de covarianzas para las estimaciones de los parámetros
- Contrastes sobre combinaciones lineales de los parámetros
- Especificación explícita de modelos anidados
- Ajuste de modelos de regresión logística condicional con emparejamiento 1-1 usando variables diferenciadas

Notas:

- Ambos de estos procedimientos ajustan un modelo para datos binarios que es un modelo lineal generalizado con una distribución binomial y la función de enlace logit. Si una función de enlace

diferente es más apropiada para sus datos, deberá utilizar el procedimiento de modelos lineales generalizados.

- Si tiene mediciones repetidas de datos binarios, o registros que están correlacionados de otra forma, deberá tener en cuenta los procedimientos de modelos mixtos lineales generalizados o ecuaciones de estimación generalizadas.

Regresión lineal Ridge

Linear Ridge utiliza la clase Python `sklearn.linear_model.Ridge` para estimar L2 o los modelos de regresión lineal regularizada por pérdida cuadrada para una variable dependiente en una o más variables independientes, e incluye modalidades opcionales para visualizar gráficos de rastreo y para seleccionar el valor de hiperparámetro alfa basado en la validación cruzada. Cuando se ajusta un único modelo o se utiliza la validación cruzada para seleccionar alfa, se puede utilizar una partición de datos reservados para estimar el rendimiento fuera de la muestra.

Además de ajustar un modelo con un valor especificado del parámetro de regularización alfa, el reborde lineal puede presentar un trazado de reborde de valores de coeficiente para un rango de valores alfa, o facilitar la elección del valor de hiperparámetro a través de la validación cruzada de k-fold en cuadrículas especificadas de valores. Si se ajusta un único modelo o se realiza una selección alfa a través de la validación cruzada, el modelo final se puede aplicar a los datos retenidos creados por una partición de los datos de entrada para obtener una estimación válida del rendimiento fuera de la muestra del modelo.

Obtención de un análisis de regresión de cresta lineal

1. En los menús seleccione:

Analizar > Regresión > Alternativas de OLS lineal > Ridge

El diálogo le permite especificar una variable que asigna cada caso del conjunto de datos activo a la muestra de entrenamiento o reservada.

2. Seleccione una variable objetivo numérica. Sólo se necesita una variable objetivo para ejecutar un análisis.
3. Especifique un dependiente numérico.
4. Especifique al menos una variable de factor categórico o una variable de covariable numérica.

Opcionalmente, **Partición** proporciona una forma de crear una reserva o un subconjunto de prueba de los datos de entrada para la estimación del rendimiento fuera de muestra del modelo especificado o elegido. Todo el particionamiento se realiza después de la supresión por lista de los casos con datos no válidos para cualquier variable utilizada por el procedimiento. Tenga en cuenta que para la validación cruzada, se crean pliegues o particiones de los datos de entrenamiento en Python. Los datos reservados creados por la partición no se utilizan en la estimación, independientemente de la modalidad en vigor.

La partición se puede definir especificando la proporción de casos asignados aleatoriamente a cada muestra (bajo **Particiones de entrenamiento y retención**), o mediante una variable que asigna cada caso a la muestra de entrenamiento o reserva. No puede especificar tanto el entrenamiento como las variables. Si no se especifica la partición, se crea una muestra reservada de aproximadamente el 30% de los datos de entrada.

El **% de entrenamiento** especifica el número relativo de casos en el conjunto de datos activo que se asignará aleatoriamente a la muestra de entrenamiento. El entrenamiento predeterminado es 70%.

Regresión de cresta lineal: Opciones

La pestaña Opciones proporciona opciones para:

Modalidad

Proporciona opciones para especificar una de las modalidades siguientes:

Ajustar con alfa especificado

Cuando se selecciona, se ajusta un único modelo a los datos de entrenamiento utilizando un único valor de regularización alfa. Este es el valor predeterminado. Si se especifica una partición, el modelo único o final ajustado se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Gráfico**, se pueden seleccionar los gráficos de los valores observados y/o de los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Gráfico de rastreo

Cuando se selecciona, se muestran tres gráficos para los datos de entrenamiento como una función de alfa para el conjunto especificado de valores de alfa:

- Un gráfico de trazado de cresta de coeficientes de regresión.
- Un gráfico de R^2 .
- Gráfico de error cuadrático medio (MSE).

Aunque se respeta la partición, no se proporcionan resultados para los datos de prueba retenidos porque no hay resultados de modelo finales de esta modalidad.

Selección alfa mediante validación cruzada

Cuando se selecciona, se realiza una búsqueda de cuadrícula con validación cruzada para evaluar modelos, y se elige el mejor alfa basándose en el mejor promedio R^2 en los pliegues de validación. El campo **Número de pliegues de validación cruzada** se puede utilizar para cambiar el valor predeterminado de 5 divisiones o pliegues para la validación cruzada. Si se especifica una partición, el modelo único o final ajustado se aplica a los datos de prueba retenidos para estimar el rendimiento fuera de muestra.

En **Visualizar**, puede seleccionar mostrar información básica sobre sólo el modelo con el valor elegido de alfa (**Mejor**), información básica sobre todos los modelos comparados (**Comparar modelos**) o información completa sobre todas las divisiones o pliegues para todos los modelos (**Comparar modelos y pliegues**). **Mejor** es el valor predeterminado.

En **Gráfico**, se pueden seleccionar gráficos de la media R^2 y/o MSE sobre pliegues de validación. También se pueden seleccionar los gráficos de los valores observados y/o los residuos frente a los valores pronosticados.

En **Guardar**, puede especificar los valores pronosticados y los residuos que desea guardar.

Especificar valores alfa individuales

Cuando se selecciona para la modalidad **Ajustar con alfa especificado**, puede especificar un único valor de regularización de alfa. Cuando se selecciona para la modalidad **Rastrear gráfico** o **Selección alfa mediante validación cruzada**, puede especificar varios valores.

Valores

Especifique uno o varios valores de regularización alfa positivos. Se pueden especificar varios valores individualmente o como rangos. El valor predeterminado es 1.

Especificar cuadrícula de valores alfa

Cuando se selecciona para la modalidad **Trazar gráfico** o **Selección alfa mediante validación cruzada**, se puede especificar una cuadrícula de valores alfa exclusivos desde un valor de **Inicio** (value1) a un valor de **Fin** (value2) con el incremento de **Por** (value3). Si se especifica, sólo se permite un conjunto válido de [value1 TO value2 BY value3]. Debe satisfacer ese $0 \leq \text{value1} \leq \text{value2} \leq 1$. En los casos en los que $\text{value1} = \text{value2}$, es equivalente a especificar un único value1, independientemente de value3.

La **métrica Alfa** para rangos de valores puede ser **Lineal** o **Base 10 logarítmica** (10 se eleva a la potencia de los valores especificados).

Los gráficos se muestran utilizando la métrica especificada para los ejes X horizontales de los valores alfa variables.

Criterios

Controla los análisis.

Incluir intersección

Incluye una interceptación en los modelos ajustados. Tenga en cuenta que el procedimiento de extensión no centra ni estandariza la variable dependiente, y la intersección no se penaliza durante la estimación.

Estandarizar predictores

Estandariza todas las variables independientes.

Número de pliegues de validación cruzada

Número de divisiones o pliegues para la evaluación de validación cruzada de modelos. Debe ser un valor entero positivo mayor que 1. El valor predeterminado es 5.

Estado aleatorio de Python

El valor de `random_state` en Python utilizado al realizar la evaluación de validación cruzada de modelos. Permite la réplica de resultados que implican números pseudoaleatorios. Debe ser un entero en el rango de 0 a $2^{32}-1$. El valor predeterminado es 0.

Límite de tiempo (minutos)

El número de minutos permitidos para que se ejecuten los cálculos del modelo. Si especifica 0, el temporizador se desactiva. El valor predeterminado es 5.

Visualización

Especifica la cantidad de salida que se va a visualizar para la modalidad **Selección alfa mediante validación cruzada**.

Mejor

Muestra sólo los resultados básicos para el mejor modelo elegido. Este es el valor predeterminado.

Comparar modelos

Muestra los resultados básicos para todos los modelos evaluados.

Comparar modelos y pliegues

Muestra resultados detallados completos para cada división o plegado para cada modelo evaluado.

Gráfico

Especifica gráficos de valores observados o residuales frente a valores pronosticados, y con validación cruzada, especificación de gráficos de promedio de error cuadrático (MSE) y/o promedio R^2 sobre pliegues de validación cruzada frente a valores alfa.

Media de error cuadrático promedio (MSE) de validación cruzada con respecto a alfa

Para la modalidad **Selección alfa mediante validación cruzada**, muestra un gráfico de líneas de MSE promedio sobre pliegues de validación cruzada frente a alfa. Para la modalidad de **Gráfico de rastreo**, se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Media de R cuadrado de validación cruzada con respecto a alfa

Para la modalidad **Selección alfa mediante validación cruzada**, muestra un gráfico de líneas de R^2 promedio sobre pliegues de validación cruzada frente a alfa. Para la modalidad de **Gráfico de rastreo**, se genera automáticamente un gráfico similar basándose en los datos de entrenamiento completos.

Observados frente a pronosticados

Muestra un diagrama de dispersión de los valores observados frente a los pronosticados para el modelo especificado o el mejor.

Residuales frente a pronósticos

Muestra un diagrama de dispersión de residuos frente a valores pronosticados para el modelo especificado o el mejor.

Guardar

Especifica variables para guardar en el conjunto de datos activo.

Valores pronosticados

Guarde los valores pronosticados del modelo especificado o mejor en el conjunto de datos activo. Opcionalmente, puede especificar un **Nombre de variable personalizada**.

Residuos

Guarde los residuos de las predicciones de modelo especificadas o mejores en el conjunto de datos activo. Opcionalmente, puede especificar un **Nombre de variable personalizada**.

Regresión logística

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

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

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

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

Consideraciones sobre datos de regresión logística

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

Supuestos. La regresión logística no se basa en supuestos distribucionales en el mismo sentido en que lo hace el análisis discriminante. Sin embargo, la solución puede ser más estable si los predictores tienen una distribución normal multivariada. Adicionalmente, al igual que con otras formas de regresión, la multicolinealidad entre los predictores puede llevar a estimaciones sesgadas y a errores estándar inflados. El procedimiento es más eficaz cuando la pertenencia a grupos es una variable categórica auténtica; si la pertenencia al grupo se basa en valores de una variable continua (por ejemplo “CI alto” en contraposición a “CI bajo”), deberá considerar el utilizar la regresión lineal para aprovechar la información mucho más rica ofrecida por la propia variable continua.

Procedimientos relacionados. Utilice el procedimiento Diagrama de dispersión para mostrar en pantalla sus datos para multicolinealidad. Si se cumplen los supuestos de normalidad multivariante y de matrices

de varianzas-covarianzas iguales, puede obtener una solución más rápida utilizando el procedimiento Análisis discriminante. Si todos los predictores son categóricos, puede además utilizar el procedimiento Loglineal. Si la variable dependiente es continua, utilice el procedimiento Regresión lineal. Puede utilizar el procedimiento Curva ROC para realizar gráficos de las probabilidades guardadas con el procedimiento Regresión logística.

Para obtener un análisis de regresión logística

1. En los menús seleccione:

Analizar > Regresión > Logística binaria...

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

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

Si lo desea, puede seleccionar casos para el análisis. Elija una variable de selección y entre los criterios de regla.

Regresión logística: Establecer regla

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

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

Métodos de selección de variables en el análisis de regresión logística

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

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

- *Eliminación hacia atrás (Wald)*. Selección por pasos sucesivos hacia atrás El contraste para la eliminación se basa en la probabilidad del estadístico de Wald.

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

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

Regresión logística: Definir variables categóricas

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

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

Covariables categóricas. Lista las variables identificadas como categóricas. Cada variable incluye una notación entre paréntesis indicando el esquema de codificación de contraste que va a utilizarse. Las variables de cadena (señaladas con el símbolo < a continuación del nombre) estarán presentes ya en la lista Covariables categóricas. Seleccione cualquier otra covariable categórica de la lista Covariables y 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 de la variable de predictor (excepto la última categoría) se compara con la categoría siguiente.
- **Polinómico.** Contrastos polinómicos ortogonales. Se supone que las categorías están espaciadas de forma equidistante. Los contrastes polinómicos sólo están disponibles para variables numéricas.
- **Desviación.** Cada categoría del predictor, excepto la categoría de referencia, se compara con el efecto global.

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

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

Regresión logística: Guardar nuevas variables

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

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

- *Probabilidades.* Para cada caso, guarda la probabilidad pronosticada de aparición del evento. En los resultados, una tabla muestra el nombre y el contenido de cualquier variable nueva. El "suceso" es la categoría de la variable dependiente con el valor superior; por ejemplo, si la variable dependiente toma los valores 0 y 1, se guarda la probabilidad pronosticada de la categoría 1.
- *Grupo de pertenencia pronosticado.* Grupo con la mayor probabilidad posterior, basado en puntuaciones discriminantes. El grupo pronosticado por el modelo al cual pertenece el caso.

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

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

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

- *Residuos no estandarizados.* Diferencia entre un valor observado y el valor predicho por el modelo.
- *Residuo logit.* El residuo del caso si se pronostica en la escala logit. El residuo logit es el residuo dividido por la probabilidad pronosticada multiplicada por 1 menos la probabilidad pronosticada.
- *Residuo estudentizado.* El cambio en la desvianza del modelo si se excluye el caso.
- *Residuos estandarizados.* 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.
- *Desvianza.* Los residuos basados en la desvianza del modelo.

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

Regresión logística: Opciones

Puede especificar estas opciones para el análisis de regresión logística:

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

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

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

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

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

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

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

Características adicionales del comando LOGISTIC REGRESSION

La sintaxis de comandos también le permite:

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

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

Regresión logística multinomial

La regresión logística multinomial es útil para situaciones en las cuales desea poder clasificar sujetos basándose en los valores de un conjunto de variables predictoras. Este tipo de regresión es similar a la regresión logística, pero más general, ya que la variable dependiente no está restringida a dos categorías.

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

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

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

Consideraciones de datos sobre la regresión logística multinomial

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

Supuestos. Se asume que la razón de las ventajas de cualquier par de categorías es independiente de las demás categorías de respuesta. Según esta suposición, por ejemplo, si se introduce un nuevo producto en un mercado, las cuotas de mercado de todos los demás productos se verán afectadas de

manera igualmente proporcional. De igual manera, dado un patrón en las covariables, se asume que las respuestas son variables multinomiales independientes.

Obtención de una regresión logística multinomial

1. En los menús seleccione:

Analizar > Regresión > Regresión Logística Multinomial...

2. Seleccione una variable dependiente.

3. Los factores son opcionales y pueden ser numéricos o categóricos.

4. Las covariables son opcionales, pero si se especifican deben ser numéricas.

Regresión logística multinomial

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

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

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

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

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

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

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

Construir términos

Para las covariables y los factores seleccionados:

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

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

Todas de 2. Crea todas las interacciones 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.

Categoría de referencia de regresión logística multinomial

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

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

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

Regresión logística multinomial: Estadísticos

Puede especificar los siguientes estadísticos para una regresión logística multinomial:

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

Modelo. Estadísticos del modelo global.

- **Pseudo R cuadrado.** Imprime el estadístico de Cox y Snell, de Nagelkerke y el R^2 McFadden.
- **Resumen de pasos.** Esta tabla resume los efectos introducidos o eliminados en cada paso, mediante un método por pasos. No se produce salvo que se especifique un modelo por pasos en el cuadro de diálogo [Modelo](#).
- **Información de ajuste de los modelos.** Esta tabla compara los modelos ajustado y de sólo intersección o nulo.
- **Criterios de información.** Esta tabla imprime tanto el criterio de información de Akaike (AIC) como el criterio de información bayesiano (BIC).
- **Probabilidades de casilla.** Imprime una tabla de las frecuencias observadas y esperadas (con los residuos) y las proporciones por patrón en las covariables y por categoría de respuesta.
- **Tabla de clasificación.** Imprime una tabla de las respuestas observadas respecto a las respuestas pronosticadas.
- **Estadísticos de bondad de ajuste de chi-cuadrado.** Imprime los estadísticos de chi-cuadrado de Pearson y de chi-cuadrado de la razón de verosimilitud. Los estadísticos se calculan para los patrones en las covariables determinados por todos los factores y las covariables o por un subconjunto de los factores y las covariables definido por el usuario.
- **Medidas de monotonía.** Muestra una tabla con información sobre el número de pares concordantes, pares discordantes y empates. La D de Somers, la gamma de Goodman y Kruskal, la tau-a de Kendall y el índice de concordancia C también se muestran en esta tabla.

Parámetros. Estadísticos relativos a los parámetros del modelo.

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

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

Regresión logística multinomial: Criterios

Puede especificar los siguientes criterios para una regresión logística multinomial:

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

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

Delta. Le permite especificar un valor no negativo menor que 1. Este valor se añade a cada celda vacía de la tabla de referencias cruzadas de la categoría de respuesta por patrón de covariable. Se ayuda así a estabilizar el algoritmo y evitar sesgos en las estimaciones.

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

Opciones de Regresión logística multinomial

Puede especificar las siguientes opciones para una regresión logística multinomial:

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

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

- **Probabilidad de entrada.** Se trata de la probabilidad del estadístico de la razón de verosimilitud para la entrada de variables. Cuanto mayor sea la probabilidad especificada, más fácil resultará que una variable entre en el modelo. Este criterio se ignora a menos que se seleccione uno de los métodos siguientes: hacia delante, pasos sucesivos hacia adelante o pasos sucesivos hacia atrás.
- **Prueba de entrada.** Éste es el método para introducir los términos en los métodos por pasos. Escoja entre la prueba de la razón de verosimilitud y la prueba de puntuación. Este criterio se ignora a menos que se seleccione uno de los métodos siguientes: hacia delante, pasos sucesivos hacia adelante o pasos sucesivos hacia atrás.
- **Probabilidad de eliminación.** Se trata de la probabilidad del estadístico de la razón de verosimilitud para la eliminación de variables. Cuanto mayor sea la probabilidad especificada, más fácil resultará que una variable permanezca en el modelo. Este criterio se ignora a no ser que se seleccione el método de Eliminación hacia atrás, Avanzar por pasos o Pasos sucesivos hacia atrás.
- **Prueba de eliminación.** Éste es el método utilizado para eliminar términos en los métodos por pasos. Puede elegir entre la prueba de la razón de verosimilitud o la prueba de Wald. Este criterio se ignora a no ser que se seleccione el método de Eliminación hacia atrás, Avanzar por pasos o Pasos sucesivos hacia atrás.
- **Efectos por pasos mínimos en el modelo.** Al utilizar los métodos de eliminación hacia atrás o pasos sucesivos hacia atrás, se especifica el mínimo número de términos que puede incluirse en el modelo. La intersección no se cuenta como término de modelo.
- **Efectos por pasos máximos en el modelo.** Al utilizar los métodos de entrada hacia adelante o Avanzar por pasos, se especifica el máximo número de términos que puede incluirse en el modelo. La intersección no se cuenta como término de modelo.

- **Restringir jerárquicamente la entrada y la eliminación de términos.** Esta opción permite elegir si se desea aplicar restricciones a la inclusión de términos de modelo. La jerarquía precisa que para que se incluya un término, todos los inferiores que formen parte del que se desea incluir se encuentren antes en el modelo. Por ejemplo, si el requisito de jerarquía está activado, los factores *Estado civil* y *Sexo* deben estar en el modelo antes de poder añadir la interacción *Estado civil*Sexo*. Las tres opciones en forma de botón radial determinan el papel de las covariables a la hora de determinar la jerarquía.

Regresión logística multinomial: Guardar

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

Variables guardadas. Se guardarán las siguientes variables:

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

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

Características adicionales del comando NOMREG

La sintaxis de comandos también le permite:

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

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

Regresión probit

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

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

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

Este procedimiento utiliza los algoritmos propuestos e implementados en NPSOL[®] por Gill, Murray, Saunders & Wright para estimar los parámetros del modelo.

Consideraciones sobre datos de la regresión probit

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

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

Procedimientos relacionados. El análisis probit está estrechamente relacionado con la regresión logística; de hecho, si elige la transformación logit, este procedimiento calculará esencialmente una regresión logística. En general, el análisis probit es apropiado para los diseños experimentales, mientras que la regresión logística es más adecuada para los estudios observacionales. Las diferencias en los resultados reflejan estas diferencias de énfasis. El procedimiento Análisis probit informa de las estimaciones de los valores efectivos para las diferentes tasas de respuesta (incluyendo la dosis efectiva para la mediana), mientras que la Regresión logística informa de las estimaciones de las razones de probabilidad (odds ratios) para las variables independientes.

Obtener un análisis de regresión probit

1. En los menús seleccione:

Analizar > Regresión > Probit...

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

Si se desea, puede seleccionarse una variable de factor. Si lo hace, utilice **Definir rango** para definir el rango de los grupos.

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

5. Seleccione el modelo **Probit** o **Logit**.

Modelo Probit

Aplica la transformación Probit (la inversa de la función acumulada de la distribución normal estándar) a las proporciones de respuesta.

Modelo Logit

Aplica la transformación logit (log de la probabilidad) a las proporciones de respuesta.

Análisis probit: Definir rango

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

Análisis probit: Opciones

Puede especificar opciones para el análisis probit:

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

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

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

Índice de respuesta natural. Permite indicar un índice de respuesta natural incluso en la ausencia del estímulo. Ninguna, Calcular a partir de los datos o Valor.

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

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

Características adicionales del comando PROBIT

La sintaxis de comandos también le permite:

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

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

Regresión cuantil

La regresión es un método estadístico que se utiliza ampliamente en modelos cuantitativos. La regresión lineal múltiple es un enfoque básico y estándar en el cual los investigadores utilizan los valores de varias variables para explicar o predecir valores de las medias de un resultado de escala. Sin embargo, en muchas circunstancias, estamos más interesados en la mediana o en un cuantil arbitrario del resultado de la escala.

La regresión cuantil modela la relación entre un conjunto de variables predictoras (independientes) y percentiles específicos (o "cuantiles") de una variable de destino (dependiente), la mayoría de las veces la mediana. Tiene dos ventajas principales sobre la regresión de los mínimos cuadrados ordinarios:

- La regresión cuantil no hace ninguna suposición sobre la distribución de la variable objetivo.
- La regresión cuantil tiende a resistir la influencia de las observaciones atípicas.

La regresión cuantil se utiliza ampliamente para la investigación en sectores como la ecología, la atención sanitaria y la economía financiera.

Ejemplo

¿Cuál es la relación entre el ingreso total de los hogares y la proporción de los ingresos que se gasta en alimentación? La ley de Engel es una observación empírica sobre economía en la que se afirma que, a medida que aumentan los ingresos, la proporción de lo que se gasta en alimentación se reduce, aunque el gasto real en alimentación en términos absolutos suba. Aplicando la regresión cuantil a estos datos, se puede determinar qué gasto de alimentación puede cubrir el 90% de las familias (para 100 familias con un ingreso determinado) cuando no buscamos saber el gasto medio en alimentación.

Estadísticos

Regresión cuantil, enfoque simplex, algoritmo de optimización no lineal de punto interior Frisch-Newton, Barrodale y Roberts, Bofinger, Hall Sheather, ancho de banda, nivel de significación, manipulaciones de matriz, criterio de convergencia, ponderaciones de regresión, término de intersección, destino pronosticado, residuo de predicción, tabulación, gráficos de predicción, estimaciones de parámetro, matriz de covarianzas, matriz de correlaciones, valores observados, intervalo de confianza.

Este procedimiento utiliza los algoritmos propuestos por Koenker, R. W. y Bassett, G. W. (1978). *Regression quantiles*, *Econometrica*, 46, 33–50.

Consideraciones de datos sobre la regresión cuantil

Datos

Es obligatoria una única variable dependiente numérica. La variable objetivo debe ser una variable continua. Los predictores pueden ser variables continuas o variables auxiliares para los predictores categóricos. Para ejecutar un análisis es obligatorio el término de intersección o como mínimo un predictor.

Supuestos

La regresión cuantil no realiza ninguna asunción sobre la distribución de la variable objetivo y resiste la influencia de observaciones atípicas.

Procedimientos relacionados

El análisis cuantil está relacionado con la regresión de mínimos cuadrados ordinarios.

Obtención de un análisis de regresión cuantil

1. En los menús seleccione:

Analizar > Regresión > Cuantil...

El diálogo le permite especificar estas variables de objetivo, factor, covariable y ponderación para utilizar para el análisis de regresión de cuantil. El diálogo también proporciona la opción de conservar la memoria para el análisis complejo o conjuntos de datos grandes.

2. Seleccione una variable objetivo numérica. Sólo se necesita una variable objetivo para ejecutar un análisis. Sólo se permiten variables numéricas.
3. Si lo desea, puede seleccionar una o más variables de factor. Las variables de escala no están permitidas.
4. Si lo desea, seleccione una o más variables covariables. Las variables de cadena no están permitidas.

Nota: Si las listas **Factores** y **Covariables** están vacías, y se ha seleccionado **Incluir la intersección en el modelo** en el diálogo Modelo, se muestra el mensaje siguiente:

```
No effects have been specified. Therefore, an intercept only model will be fit.
Do you want to fit an intercept-only model?
```

5. Opcionalmente, seleccione una variable de ponderación de regresión. Las variables de cadena no están permitidas.
6. De forma opcional, seleccione **Conservar memoria para análisis complejos o conjuntos de datos grandes**. Este valor controla si los datos se mantienen o no en un archivo externo durante el procesamiento. La habilitación de este valor puede ayudar a conservar los recursos de memoria al ejecutar análisis complejos, o análisis con conjuntos de datos grandes.

Regresión cuantil: Criterios

El diálogo Criterios proporciona las siguientes opciones:

Cuantil

Proporciona opciones para especificar los cuantiles.

Especificar cuantiles únicos

Cuando está seleccionado, se requiere al menos un valor para ejecutar el análisis. Están permitidos varios valores y cada valor debe pertenecer a $[0, 1]$. Puede especificar varios valores separando cada valor con un espacio en blanco (o espacios en blanco). Utilice los botones **Añadir**, **Cambiar** y **Eliminar** para gestionar los valores en la lista de valores de cuantil.

Todos los valores deben ser exclusivos (los valores duplicados no están permitidos). El valor predeterminado es 0.5.

Especificar cuantiles de cuadrícula

Cuando está seleccionado, puede especificarse una cuadrícula de cuantiles desde un valor **Inicial** (valor1) hasta un valor **Final** (valor2) con el incremento de **Por** (valor3). Si se especifica, solo se permite utilizar un conjunto válido de $[\text{valor1 TO valor2 BY valor3}]$. Debe ajustarse al esquema $0 \leq \text{valor1} \leq \text{valor2} \leq 1$. En casos en los que $\text{valor1} = \text{valor2}$, es el equivalente a especificar un solo valor1, independientemente del valor3.

Método de estimación

Proporciona opciones para especificar el método de estimación del modelo.

Elegido automáticamente por el programa

Permite que el procedimiento seleccione automáticamente el método de estimación adecuado. Ésta es el ajuste predeterminado.

Algoritmo simplex

Llama al algoritmo simplex desarrollado por Barrodale y Roberts.

Optimización no lineal de punto interior Frisch-Newton

Llamadas al algoritmo de optimización no lineal de punto interior Frisch-Newton.

Post-estimación

Proporciona opciones para la post-estimación de la varianza-covarianza de las estimaciones de los parámetros y los intervalos de confianza para los valores de destino pronosticados.

Asumir casos son IID

Cuando se selecciona, este valor asume que los términos de error se distribuyan de forma independiente e idéntica (IDD). Cuando el valor no está seleccionado, el tiempo de cálculo puede aumentar significativamente para los modelos de gran tamaño. Éste es el valor seleccionado de forma predeterminada.

Tipo de ancho de banda

Determina qué método de ancho de banda se utiliza para estimar la matriz de varianza-covarianza de las estimaciones de los parámetros (**Bofinger** o **Hall-Sheather**). **Bofinger** es el valor predeterminado.

Método numérico

Proporciona las opciones siguientes:

Tolerancia para la singularidad

Especifica el valor de tolerancia para las manipulaciones de matriz en el método de punto interior. El valor especificado debe ser un único valor doble en $(0, 10^{-3})$, con 10^{-12} como valor predeterminado.

Convergencia

Especifica el criterio de convergencia para el método numérico. El valor especificado debe ser un único valor doble en $(0, 10^{-3})$, con 10^{-6} como valor predeterminado.

Número máximo de iteraciones

Especifica el número máximo de iteraciones. El valor especificado debe ser un único entero positivo. El valor predeterminado es 2000.

Valores perdidos

Proporciona opciones para determinar cómo se manejan los valores perdidos.

Excluir tanto los valores perdidos del usuario como los valores perdidos del sistema

Cuando se selecciona, se excluyen ambos valores, los perdidos del usuario y los perdidos del sistema.

Los valores perdidos del usuario se tratan como válidos

Cuando está seleccionado, los valores perdidos del usuario se tratan como válidos.

Intervalos de confianza (%)

Especifica el nivel de significación. Cuando se especifica, el valor debe ser un único valor doble entre 0 y 100. El valor predeterminado es 95.

Regresión cuantil: Modelo

El diálogo Modelo proporciona opciones para especificar los efectos y las ponderaciones que se utilizan en el modelo. Si se omite, o se especifica por sí mismo, el modelo contendrá el término de intersección y todos los efectos principales con las covariables de la lista de covariables y los factores de la lista de factores.

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. Si selecciona **Construir términos**, estarán disponibles las siguientes opciones de efectos e interacción para términos no anidados.

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.

Si selecciona **Construir términos anidados**, puede crear términos anidados. 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.

Notas sobre anidamiento:

- Para incluir un efecto para una interacción entre dos factores, utilice la palabra clave BY o el asterisco (*) para unir los factores involucrados en la interacción.
- Los factores dentro de un efecto de interacción deben ser distintos.
- Utilice pares de paréntesis para incluir un efecto de anidamiento de un término en otro.
- Cuando estén presentes más de un par de paréntesis, cada par de paréntesis debe estar dentro o anidado dentro de otro par de paréntesis.

- Se permiten varias anidaciones.
- No se admiten las interacciones entre los efectos anidados.

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 $A(X)$.

Incluir la interceptación en el modelo

Cuando se selecciona, el término de intersección se incluye en el modelo. Cuando no está seleccionado, se requiere al menos un predictor para ejecutar el análisis. Esta opción está habilitada de forma predeterminada.

Regresión cuantil: Visualización

El diálogo Visualización proporciona valores de resultado y de control de representación.

Imprimir

Están disponibles las opciones de resultado siguientes.

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.

Representar y tabular

Están disponibles las opciones de representación siguientes:

Representar las estimaciones de parámetro para

Puede seleccionar representar las estimaciones de parámetros para un número específico de efectos principales, o para todos los efectos del modelo. El valor **xx efectos principales** controla el número de categorías, o combinaciones de categoría en un efecto mixto, que se trazan en la interacción de una covariable y uno o dos factores. El valor debe ser un único entero positivo (50 es el valor predeterminado).

Notas:

- Los gráficos de predicción se crean para todos los efectos cuando el valor entero que especifique sea mayor que el número de categorías o combinaciones.
- El valor sólo es válido cuando se especifican varios valores para el valor **Valores de cuantil** en el diálogo Criterios. No se crean gráficos cuando se especifica un único valor de cuantil.

Mostrar lo pronosticado por gráfico observado

Controla la creación del gráfico de valores pronosticados frente a los observados. Si está habilitado, se crea un solo gráfico que contiene los puntos (con diferentes puntos coloreados que representan diferentes cuantiles). De forma predeterminada, el valor está inhabilitado.

Pronosticar los efectos del modelo

Si está habilitado, están disponibles las opciones siguientes:

Representar o tabular los x efectos principales

Especifique el número de efectos principales cuyo gráfico de predicción o tabla de predicción se creará. 3 es el valor predeterminado.

Nota: Los gráficos de predicción o las tablas de predicción se crean para todos los efectos cuando el valor especificado es mayor que el número de efectos válidos del modelo.

Representar o tabular efectos especificados por el usuario

Las directrices de efectos válidos son:

- Efecto con una covariable (incluida una potencia alta de la misma covariable): Cree un único gráfico que contenga las líneas pronosticadas mediante distintos cuantiles.
- Efecto con un factor: Tabule las predicciones para las categorías del factor mediante diferentes cuantiles.
- Efecto con la interacción de dos factores: Para cada cuantil, tabule las predicciones para las categorías de dos factores.
- Efecto con la interacción de una covariable y uno o dos factores: Para cada cuantil, cree un gráfico que contenga las líneas para cada categoría o combinación de las categorías dentro del efecto de interacción.
- El número máximo de combinaciones que se va a representar se controla mediante el valor especificado en **Representar máximo xx categorías o combinaciones de categorías en un efecto mixto**.

Los efectos que se mueven de la lista **Efectos del modelo** a la lista **Líneas de predicción** se utilizan para la representación. Los gráficos no se crean en los casos en los que los efectos especificados son constantes (se eliminan de la generación de modelos).

Los efectos que se mueven de la lista **Efectos del modelo** a la lista **Tablas de predicción** se utilizan para la tabulación. Las tablas no se crean en los casos en los que los efectos especificados son constantes (se eliminan de la generación de modelos).

Representar máximo xx categorías o combinaciones de categorías en un efecto mixto

Controla el número máximo de combinaciones de categorías que se va a representar. El valor predeterminado es 10.

Regresión cuantil: Guardar

El diálogo Guardar proporciona opciones para puntuar el modelo.

Valor pronosticado de respuesta

Cuando está seleccionado, se puntúa el valor de destino pronosticado.

Residual

Cuando se selecciona, se puntúan los residuos de predicción.

Límite inferior del intervalo de predicción

Cuando se selecciona, se puntúan los límites inferiores de los intervalos de predicción.

Límite superior del intervalo de predicción

Cuando está seleccionado, se puntúan los límites superiores de los intervalos de predicción.

Nota: Se puede especificar un nombre de variable para cada opción de guardar. Si se especifica un nombre raíz, este debe ser un nombre de variable válido. Un nombre raíz, seguido de un carácter de subrayado "_" y un sufijo de cuantil significativo, se utiliza cuando se especifican varios valores en el valor **Valores de cuantil** en el diálogo Criterios.

Regresión cuantil: Exportar

El diálogo Exportar proporciona opciones para especificar qué estadísticos se exportan, cómo se exportan los estadísticos (archivo de datos externo o conjuntos de datos), y controlar cómo se manejan los datos durante el proceso (proceso normal o retenido en un archivo de trabajo externo mientras se procesa).

Matriz de covarianzas de estimaciones de parámetros

Cuando se selecciona, se habilitan las opciones para escribir la matriz de covarianza de las estimaciones de los parámetros en un archivo de datos externo, o en un conjunto de datos previamente declarado.

Matriz de correlaciones de estimaciones de parámetros

Cuando se selecciona, se habilitan las opciones para escribir la matriz de correlaciones de las estimaciones de los parámetros en un archivo de datos externo o en conjunto de datos declarado previamente.

La matriz de covarianza/correlación se guardará en un único conjunto de datos o en un archivo externo en la presencia de varios cuantiles de regresión.

Ante la presencia de varios cuantiles, esta opción conmuta el almacenamiento de las matrices de covarianzas/correlaciones a uno o varios conjuntos de datos o archivos de datos externos. Cuando no está habilitado, las matrices se guardan en un único archivo de datos externo o en un conjunto de datos. Cuando está habilitado, las matrices se guardan en varios archivos de datos externos o conjuntos de datos. El valor entra en vigor sólo cuando se especifican varios valores para el valor **Valores de cuantil** en el diálogo Criterios.

Nota: Esta opción sólo está disponible si se ha seleccionado **Matriz de covarianzas de estimaciones de parámetros** o **Matriz de correlaciones de estimaciones de parámetros**.

Exportar información del modelo a un archivo XML

Cuando está seleccionado, proporciona opciones para exportar la información del modelo a un nombre y una ubicación de archivo XML específicos.

Exportar como XML

Cuando está seleccionado **Exportar información de modelo a archivo XML**, puede seleccionar exportar estimaciones de parámetro y matrices de covarianza o solo estimaciones de parámetro. El valor predeterminado es **Estimaciones de los parámetros y matriz de covarianzas**.

Convenciones de nombres de archivos

- Cuando se especifica un valor único para el valor **Valores de cuantil** en el diálogo Criterios, se utiliza `savefile` y `dataset` para dar nombre al archivo de datos externo o conjunto de datos.
- Cuando se especifican varios valores para el valor **Valores de cuantil** en el diálogo Criterios, cada cuantil se guarda en un archivo de datos externo o conjunto de datos.
- Un carácter de subrayado "_", seguido de un sufijo cuantil significativo, se añade automáticamente al nombre del archivo de datos o del conjunto de datos. Por ejemplo, si se especifica 0.25, 0.50 y 0.75 como **Valores de cuantil**, se añaden los sufijos `_25`, `_50` y `_75` a los nombres de archivos de datos (antes de la extensión `.sav`).
- Se pueden especificar dígitos adicionales para cada sufijo de cuantil (si es necesario).
- El cero inicial y el punto decimal de **Valores de cuantil** no se utilizan en el sufijo.
- Cuando se especifica una notación científica en **Valores de cuantil**, se convierte a valor decimal cuando se visualiza en el sufijo.

Regresión no lineal

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

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

Estadísticos. Para las iteraciones: estimaciones de los parámetros y suma de cuadrados residual. Para los modelos: suma de cuadrados para regresión, residual, total corregido y no corregido, estimaciones de los parámetros, errores estándar asintóticos y matriz de correlaciones asintóticas de estimaciones de los parámetros.

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

Consideraciones sobre datos de regresión no lineal

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

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

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

Para obtener un análisis de regresión no lineal

1. En los menús, seleccione:

Analizar > Regresión > No lineal...

2. Seleccione una variable numérica dependiente de la lista de variables del conjunto de datos activo.

3. Para crear una expresión de modelo, introduzca la expresión en el campo **Expresión de modelo** o pegue componentes (variables, parámetros, funciones) en el campo.

4. Identifique los parámetros del modelo pulsando en **Parámetros**.

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

Lógica condicional (Regresión no lineal)

Se puede especificar un modelo segmentado utilizando la lógica condicional. Para usar la lógica condicional dentro de la expresión del modelo o de la función de pérdida, debe generar la suma de una serie de términos, uno para cada condición. Cada término se compone de una expresión lógica (entre paréntesis) multiplicada por la expresión que resultará cuando esa expresión lógica es verdadera.

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

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

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

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

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

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

Se pueden generar con facilidad ejemplos más complicados reemplazando diferentes expresiones lógicas y expresiones de resultado. Recuerde que las desigualdades dobles, como $0 < X < 1$, must be written as compound expressions, such as $(X > 0 \ \& \ X < 1)$.

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

(ciudad='Madrid')*costliv + (ciudad='Guadalajara')*0.59*costliv

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

Regresión no lineal: Parámetros

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

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

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

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

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

Modelos comunes de regresión no lineal

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

Tabla 1. Ejemplos de sintaxis de modelo

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

Tabla 1. Ejemplos de sintaxis de modelo (continuación)

Nombre	Expresión del modelo
Razón de cuadráticas	$(b1 + b2 * x + b3 * x **2) / (b4 * x **2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x ** b4)$
Densidad de rendimiento	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

Regresión no lineal: Función de pérdida

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

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

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

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

Regresión no lineal: Restricciones para los parámetros

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

Cada ecuación o desigualdad requiere los siguientes elementos:

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

Regresión no lineal: Guardar variables nuevas

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

- *Residuos*. Guarda los residuos con el nombre de variable resid.

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

Regresión no lineal: Opciones

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

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

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

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

Regresión no lineal: Interpretación de los resultados

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

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

Características adicionales del comando NLR

La sintaxis de comandos también le permite:

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

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

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

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

Estimación ponderada

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

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

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

Consideraciones sobre datos de estimación ponderada

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

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

Procedimientos relacionados. El procedimiento Explorar se puede utilizar para inspeccionar los datos. Este procedimiento proporciona pruebas de normalidad y homogeneidad de la varianza, así como representaciones gráficas. Si la variable dependiente parece tener la misma varianza para todos los

niveles de las variables independientes, puede utilizar el procedimiento Regresión lineal. Si los datos parecen violar un supuesto (como puede ser la normalidad), intente transformarlos. Si los datos no están relacionados linealmente y una transformación no ayuda, utilice un modelo alternativo en el procedimiento Estimación curvilínea. Si la variable dependiente es dicotómica (por ejemplo, si se ha completado una determinada venta o si el elemento es defectuoso) utilice el procedimiento Regresión logística. Si su variable dependiente está censurada (por ejemplo, el tiempo de supervivencia después de la cirugía), use Tablas de Vida, Kaplan-Meier o Regresión de Cox, disponibles en Tablas Personalizadas y Estadísticas Avanzadas. Si los datos no son independientes (por ejemplo, si observa a la misma persona en diversas condiciones), utilice el procedimiento Medidas repetidas disponible en Tablas personalizadas y Estadísticas avanzadas.

Para obtener un análisis de Estimación ponderada

1. En los menús, seleccione:

Analizar > Regresión > Estimación ponderada...

2. Seleccione una variable dependiente.

3. Seleccionar una o más variables independientes.

4. Seleccione la variable que sea el origen de heterocedasticidad como variable de ponderación.

Variable de ponderación

Los datos son ponderados por el inverso de esta variable elevada a una potencia. La ecuación de regresión se calcula para cada uno de los valores del rango de potencias especificado y se indica la potencia que maximiza el logaritmo de la función de verosimilitud.

Rango de potencia

Se utiliza en conjunción con la variable de ponderación para calcular las ponderaciones. Se ajustarán varias ecuaciones de regresión, una para cada valor del rango de potencias. Los valores introducidos en el cuadro Rango de potencias y en el cuadro Hasta, deben encontrarse entre -6,5 y 7,5, ambos inclusive. Los valores de la potencia irán del valor menor al mayor, con el incremento determinado por el valor especificado. El número total de valores del rango de potencias está limitado a 150.

Estimación ponderada: Opciones

Puede especificar las opciones para el análisis de estimación ponderada:

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

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

Características adicionales del comando WLS

La sintaxis de comandos también le permite:

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

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

Regresión de cuadrados mínimos de dos etapas

Los modelos de regresión lineal estándar asumen que los errores de la variable dependiente no están correlacionados con la variable o variables independientes. Cuando éste no es el caso (por ejemplo, cuando las relaciones entre las variables son bidimensionales), la regresión lineal mediante mínimos cuadrados ordinarios (OLS) deja de proporcionar estimaciones óptimas del modelo. La regresión por

mínimos cuadrados en dos fases utiliza variables instrumentales que no estén correlacionadas con los términos de error para calcular los valores estimados de los predictores problemáticos (en la primera fase) y después utiliza dichos valores calculados para estimar un modelo de regresión lineal para la variable dependiente (la segunda fase). Dado que los valores calculados se basan en variables que no están correlacionadas con los errores, los resultados del modelo en dos fases son óptimos.

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

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

Consideraciones sobre datos de Regresión por mínimos cuadrados en dos fases

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

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

Procedimientos relacionados. Si piensa que ninguna de las variables predictoras está correlacionada con los errores de la variable dependiente, puede utilizar el procedimiento Regresión lineal. Si los datos parecen violar alguno de los supuestos (como la normalidad o la varianza constante), pruebe a transformarlos. Si los datos no están relacionados linealmente y una transformación no ayuda, utilice un modelo alternativo en el procedimiento Estimación curvilínea. Si la variable dependiente es dicotómica, por ejemplo si se ha completado o no una determinada venta, utilice el procedimiento Regresión logística. Si los datos no son independientes (por ejemplo, si observa a la misma persona en diversas condiciones), utilice el procedimiento Medidas repetidas.

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

1. En los menús seleccione:

Analizar > Regresión > Cuadrados mínimos de dos etapas...

2. Seleccione una variable dependiente.

3. Seleccione una o más variables (predictoras) explicativas.

4. Seleccione una o más variables instrumentales.

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

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

Regresión por mínimos cuadrados en dos fases: Opciones

Puede seleccionar las opciones siguientes para su análisis:

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

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

Características adicionales del comando 2SLS

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

Esquemas de codificación de variables categóricas

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

Desviación

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

```
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)  -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.

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 34. 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 COEFO = 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 34.
- 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**.

Modelos de tiempo de anomalía acelerada paramétrica

Un análisis de modelo AFT (Tiempo de anomalía acelerada paramétrica) invoca el procedimiento de modelos de supervivencia paramétrica con datos de tiempo de vida no recurrentes. Los modelos paramétricos de supervivencia asumen que el tiempo de supervivencia sigue una distribución conocida, y este análisis se ajusta a los modelos de tiempo de fallo acelerado con sus efectos de modelo proporcionales con respecto al tiempo de supervivencia.

Obtención de un análisis de modelos de tiempo de anomalía acelerada paramétrica

1. En los menús seleccione:

Analizar > Supervivencia > Modelos de Tiempo de anomalía acelerada paramétrica (AFT)

2. Seleccione una variable de origen.

Hora**Supervivencia**

Variable numérica única que indica la duración del tiempo de supervivencia.

Inicio/Fin

Variables numéricas que indican **Hora de inicio** y **Hora de finalización**.

estado

Serie opcional única o variable numérica que determina uno de los siguientes valores de estado:

Error/Suceso

Correlaciona un registro con una categoría de anomalía/suceso. El valor predeterminado para una variable de estado de serie es F.

Censura derecha

Correlaciona un registro con una categoría de censura derecha. El valor predeterminado para una variable de estado de cadena es R.

Censura izquierda

Correlaciona un registro con una categoría de censura izquierda. El valor predeterminado para una variable de estado de serie es L.

Censura de intervalo

Correlaciona un registro con una categoría de censura de intervalo. Sólo para **Inicio/Fin**. El valor predeterminado para una variable de estado de cadena es I.

Tratamiento de valores no correlacionados

Controla con qué categoría correlacionar los registros no correlacionados. Para suprimir los registros que no se han podido correlacionar, seleccione **Excluirlos del análisis**.

Para **Survival**, el estado predeterminado para todos los casos es **Failure/Event**. Para **Inicio/Fin**, el estado predeterminado es **Censura de intervalo**. Pulse el botón **Definir suceso** para definir un suceso para la variable de estado.

Covariables

Una o más variables numéricas opcionales que se tratarán como covariables. Tenga en cuenta que una variable no se puede especificar mediante **Covariables** y **Factor (s) fijo (s)**.

Factores fijos:

Una o más variables opcionales que se van a tratar como factores. Una variable no se puede especificar mediante **Factor (s) fijo (s)** y **Covariable (s)**.

Truncamiento izquierdo

Variable numérica opcional única para el truncamiento a la izquierda sólo para **Supervivencia**.

Modelos de tiempo de anomalía acelerada paramétrica: Criterios

Criterios

Un panel opcional para especificar los criterios generales.

intervalo de confianza

Un porcentaje opcional para especificar el nivel para los intervalos de confianza de los parámetros de regresión. Debe ser un único valor numérico entre 0 y 100. El valor predeterminado es 95%.

Valores perdidos

Una opción para controlar cómo se tratan los valores perdidos del usuario:

Excluir tanto los valores perdidos del usuario como los valores perdidos del sistema

Trata los valores perdidos del usuario como valores válidos. Este es el valor predeterminado.

Los valores perdidos del usuario se tratan como válidos

Ignora las designaciones de valores perdidos del usuario y las trata como valores válidos.

Tratamiento del estado

Sólo para **Inicio/Fin**. Una opción para controlar cómo tratar los registros con campos de estado incorrectos:

Descartar registro en conflicto

Descarta los registros en conflicto. Este es el valor predeterminado.

Obtener la información de tiempo según el estado

Obtiene la información de tiempo según el estado.

Derivar el estado según la información de tiempo

Cambia el estado según la información de hora.

Modelos de tiempo de anomalía acelerada paramétrica: Modelo

Modelo

Un panel opcional para especificar las opciones y valores del modelo.

Distribución del tiempo de supervivencia

Opción para especificar la distribución del tiempo de supervivencia.

Weibull

Especifica la distribución de Weibull. Este es el valor predeterminado.

Exponencial

Especifica la distribución exponencial.

Log-normal

Especifica la distribución log-normal.

Log-logística

Especifica la distribución logístico-logarítmica.

Valores de Covariables

Especifique variables de covariable.

Valores de factores

Especifique variables de factor.

Valor inicial de Intercept

Opción para especificar el valor inicial del término de interceptación. Si se especifica, debe ser un único valor numérico y no puede ser 0.

Parámetro de valor inicial de escala

Una opción para controlar el valor del parámetro de escala.

Error estándar de la regresión de OLS correspondiente

Utiliza el error estándar de la regresión de mínimos cuadrados ordinarios correspondiente como valor inicial.

Invierte el error estándar de la regresión de OLS correspondiente

Utiliza el recíproco del error estándar.

Valor suministrado por el usuario

Si se especifica un único valor numérico, el valor se utiliza como valor inicial. Si se especifica, debe ser mayor que 0.

Modelos de tiempo de anomalía acelerada paramétrica: Estimación

Estimación

Un panel opcional para especificar los valores para controlar la estimación de los modelos de tiempo de anomalía acelerada y el proceso de selección de características opcional.

Método de dirección alternativa o multiplicadores (ADMM)**Rápido**

Aplica el método de dirección alternativa rápida de multiplicadores (ADMM). Este es el valor predeterminado.

Tradicional

Aplica el algoritmo ADMM tradicional.

Aplicar la regularización L-1

Lleva a cabo el proceso para controlar la selección de características. El campo **Parámetro de penalización** especifica el parámetro de penalización que controla el proceso de regularización. Debe ser un valor único mayor que 0. El valor predeterminado es 0,001.

Criterios de convergencia de modelos**Convergencia de los parámetros**

Especifica los criterios de convergencia para el parámetro. Debe ser un único valor numérico que pertenezca a $[0, 1)$. El valor predeterminado es 0,000001. Para **Tipo**, puede seleccionar **ABSOLUTE** para aplicar la convergencia absoluta a la optimización interna o **RELATIVE** para aplicar la convergencia relativa a la optimización interna. El campo **Valor** opcional especifica una palabra clave.

Convergencia de funciones objetivas

Especifica los criterios de convergencia para la función objetivo. Debe ser un único valor numérico que pertenezca a $[0, 1)$. El valor predeterminado es 0, que no aplica los criterios de convergencia. Para **Tipo**, puede seleccionar **ABSOLUTE** para aplicar la convergencia absoluta a la optimización interna o **RELATIVE** para aplicar la convergencia relativa a la optimización interna. El campo **Valor** opcional especifica una palabra clave.

Convergencia hessiana

Especifica los criterios de convergencia para la matriz hessiana. Debe ser un único valor numérico que pertenezca a $[0, 1)$. El valor predeterminado es 0, que no aplica los criterios de convergencia. Para **Tipo**, puede seleccionar **ABSOLUTE** para aplicar la convergencia absoluta a la optimización interna o **RELATIVE** para aplicar la convergencia relativa a la optimización interna. El campo **Valor** opcional especifica una palabra clave.

Criterios de convergencia residuales

Opción para controlar el proceso de optimización.

Tanto residual primario como residual dual

Aplica tanto el criterio de convergencia residual primario como el doble. Este es el valor predeterminado.

Solo residual primario

Aplica el criterio de convergencia residual primario.

Solo residual dual

Aplica el criterio de convergencia residual dual.

Método

Parámetro opcional para especificar el método de estimación.

Auto

Elige automáticamente el método basado en el conjunto de datos de ejemplo. Este es el valor predeterminado. El campo **Número de umbral de predictores** especifica el umbral del número de predictores y debe ser un entero único mayor que 1. El valor predeterminado es 1000.

Newton-Raphson

Aplica el método de Newton-Raphson.

L-BFGS

Aplica el algoritmo BFGS de memoria limitada. El campo **Actualizar** especifica el número de actualizaciones pasadas mantenidas por el algoritmo BFGS de memoria limitada y debe ser un entero único mayor o igual que 1. El valor predeterminado es 5.

Iteración

Número máximo de iteraciones

Especifica el número máximo de iteraciones. Debe ser un entero único que pertenezca a $[1, 100]$. El valor predeterminado es 20.

Máxima subdivisión por pasos

Especifica el número máximo de subdivisión por pasos. Debe ser un entero único que pertenezca a $[1, 20]$. El valor predeterminado es 5.

Número máximo de búsquedas de línea

Especifica el número máximo de búsquedas de línea. Debe ser un entero único que pertenezca a $[1, 100]$. El valor predeterminado es 20.

Convergencia absoluta para el proceso de iteración

Especifica la convergencia absoluta para el proceso de iteración externa. Debe ser un único valor numérico que pertenezca a $(0, 1)$. El valor predeterminado es 0,0001.

Convergencia relativa para el proceso de iteración

Especifica la convergencia relativa para el proceso de iteración externa. Debe ser un único valor numérico que pertenezca a $(0, 1)$. El valor predeterminado es 0.01.

Modelos de tiempo de anomalía acelerada paramétrica: Imprimir

Imprimir

Un panel opcional para controlar las salidas de tabla.

Detalles de codificación de factores

Si se selecciona, muestra e imprime los detalles de codificación de los factores. El proceso se ignora si no hay factores en vigor.

Valores iniciales asignados a los parámetros de regresión

Si se selecciona, muestra los valores iniciales utilizados en el proceso de estimación.

Historial de iteraciones de modelo

Si se selecciona, muestra el historial de iteraciones del análisis de supervivencia. En el campo **Número de pasos**, especifique el número de pasos entre 1 y 99999999. El valor predeterminado es 1.

Resultados de selección que contienen

Controla la visualización de los detalles de la selección de características.

Variables seleccionadas y no seleccionadas

Mostrar las variables seleccionadas y no seleccionadas en la tabla.

Sólo variables seleccionadas

Mostrar sólo las variables seleccionadas.

Sólo variables no seleccionadas

Mostrar sólo las variables no seleccionadas. El campo **Número máximo de variables a visualizar** especifica el número máximo de variables que se imprimen en la tabla. El valor predeterminado es 30.

Modelos de tiempo de anomalía acelerada paramétrica: predecir

Pronosticar

Un panel opcional para puntuar y guardar las estadísticas pronosticadas en el conjunto de datos activo.

Valores de tiempo para puntuación

Valores de tiempo definidos por variables dependientes

Puntúa las **Predicciones** basándose en la variable de tiempo especificada para el modelo de supervivencia paramétrico.

Intervalos regulares

Puntúa las **Predicciones** basándose en valores de tiempo futuros. El campo **Intervalo de tiempo** especifica el intervalo de tiempo y debe ser un único valor numérico mayor que 0. El campo **Número de periodos de tiempo** especifica el número de periodos de tiempo y debe ser un entero numérico único entre 2 y 100.

Duración

Puntúa las **Predicciones** en función de la duración del tiempo para definir los valores de tiempo futuros. Debe ser una única variable numérica.

Predicciones

Supervivencia

Puntúa y guarda las estadísticas de supervivencia pronosticadas en el conjunto de datos activo. El nombre de variable personalizada predeterminado (o nombre raíz) es `PredSurvival`.

Riesgo

Puntúa y guarda los riesgos previstos en el conjunto de datos activo. El nombre de variable personalizada predeterminado (o nombre raíz) es `PredHazard`.

Riesgo acumulado

Puntúa y guarda los riesgos acumulados pronosticados en el conjunto de datos activo. El nombre de variable personalizada predeterminado (o nombre raíz) es `PredCumHazard`.

Supervivencia condicional

Puntúa y guarda las estadísticas de supervivencia condicional pronosticadas en el conjunto de datos activo. El nombre de variable personalizada predeterminado (o nombre raíz) es `PredConditionalSurvival`. El proceso se ignorará si no se especifica `PASTTIME`. Se necesita

un valor de **Tiempo de supervivencia pasado** y especifica los valores de tiempo pasado para la puntuación. Debe ser una única variable numérica.

Modelos de tiempo de anomalía acelerada paramétrica: Gráfico

Gráfico

Gráfica residual de Cox-Snell

Seleccione **Mostrar el gráfico** para crear un gráfico residual de Cox-Snell. En el campo **Número de puntos de corte agrupados**, especifique un número de 1 a 10000. El valor predeterminado es 100.

Gráficas de función

Una opción para controlar los gráficos de función.

Tipo

Supervivencia

Crea el gráfico para las funciones de supervivencia.

Riesgo

Crea el gráfico para las funciones de riesgo.

Densidad

Crea un gráfico para las funciones de densidad.

Número de puntos a visualizar

Especifica el número de puntos de función entre 1 y 200. El valor predeterminado es 100.

Valores de covariables para gráfica

Opcional para especificar los valores proporcionados por el usuario y asignarlos a los predictores. De forma predeterminada, los gráficos designados se crearán en la **Media** de cada covariable en vigor y la frecuencia de categoría de cada factor en vigor. Si se especifica, los gráficos designados se crearán basándose en el valor del patrón. En presencia de cualquier variable duplicada, se reconocería la especificada en primer lugar y el resto se ignoraría. Una variable válida debe estar contenida en un efecto de modelo. Para una covariable, el valor proporcionado por el usuario debe ser numérico. La omisión de una variable en vigor indica que la frecuencia de la categoría y la **Media** se utilizarían de forma predeterminada para el factor y la covariable, respectivamente. Si se asigna un valor no válido a una variable, el patrón solicitado no se trazará.

Valores de factores para gráfica

Opcional para especificar los valores proporcionados por el usuario y asignarlos a los predictores. En presencia de cualquier variable duplicada, se reconocería la especificada en primer lugar y el resto se ignoraría. Una variable válida debe estar contenida en un efecto de modelo. La omisión de una variable en efecto indica que la frecuencia de la categoría y la media se utilizarían por defecto para el factor y la covariable, respectivamente. Si se asigna un valor no válido a una variable, el patrón solicitado no se trazará.

Líneas separadas para

Opción para especificar una variable categórica mediante la cual se trazarán los gráficos de líneas.

Número máximo de líneas en un gráfico

Especifica el número máximo de líneas de un gráfico si se especifica **Separar líneas para**. El valor predeterminado es 10.

Modelos de tiempo de anomalía acelerada paramétrica: Exportar

Exportar

Seleccione **Exportar información de modelo a archivo XML** para escribir la información de modelo y parámetro en un archivo PMML para la puntuación. Debe especificar el directorio y el nombre de archivo del archivo PMML que se va a guardar.

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 regresión logística [5](#)
análisis de supervivencia
 en Regresión Ridge de kernel [32](#)
Análisis probit
 características adicionales del mandato [15](#)
 criterios [14](#)
 definir rango [14](#)
 estadísticos [14](#)
 índice de respuesta natural [14](#)
 intervalos de confianza fiduciaria [14](#)
 iteraciones [14](#)
 potencia relativa de la mediana [14](#)
 prueba de paralelismo [14](#)

B

bondad de ajuste
 en Regresión logística multinomial [11](#)

C

casillas con 0 observaciones
 en Regresión logística multinomial [12](#)
categoría de referencia
 en Regresión logística multinomial [11](#)
chi-cuadrado de Pearson
 bondad de ajuste [11](#)
Chi-cuadrado de Pearson
 estimación del valor del escalamiento para la dispersión
 [12](#)
clasificación
 en Regresión logística multinomial [9](#)
contrastes
 en Regresión logística [7](#)
covariables
 en Regresión logística [7](#)
covariables categóricas [7](#)
covariables de cadena
 en Regresión logística [7](#)
criterio de convergencia
 en Regresión logística multinomial [12](#)

D

D de Cook
 en Regresión logística [7](#)
delta
 corrección para casillas con 0 observaciones [12](#)
DfBeta
 en Regresión logística [7](#)

E

eliminación hacia atrás

eliminación hacia atrás (*continuación*)
 en Regresión logística [6](#)
estadístico de bondad de ajuste de Hosmer-Lemeshow
 en Regresión logística [8](#)
Estimación ponderada
 almacenamiento de la mejor ponderación como una
 variable nueva [27](#)
 características adicionales del mandato [27](#)
 ejemplo [26](#)
 estadísticos [26](#)
 historial de iteraciones [27](#)
 log-verosimilitud [26](#)
 mostrar ANOVA y estimaciones [27](#)
estimaciones de parámetros
 en Regresión logística multinomial [11](#)

F

función de desviación
 estimación del valor del escalamiento para la dispersión
 [12](#)

H

historial de iteraciones
 en Regresión logística multinomial [12](#)

I

intercept
 incluir o excluir [10](#)
intervalos de confianza
 en Regresión logística multinomial [11](#)
intervalos de confianza fiduciaria
 en el análisis probit [14](#)
iteraciones
 en el análisis probit [14](#)
 en Regresión logística [8](#)
 en Regresión logística multinomial [12](#)

L

ley de Metcherlich de devoluciones decrecientes
 en Regresión no lineal [23](#)
log-verosimilitud
 en Estimación ponderada [26](#)
 en Regresión logística multinomial [11](#)

M

matriz de correlaciones
 en Regresión logística multinomial [11](#)
matriz de covarianzas
 en Regresión logística multinomial [11](#)
modelo de densidad
 en Regresión no lineal [23](#)

- modelo de densidad de rendimiento
 - en Regresión no lineal [23](#)
- modelo de Gauss
 - en Regresión no lineal [23](#)
- modelo de Gompertz
 - en Regresión no lineal [23](#)
- modelo de Johnson-Schumacher
 - en Regresión no lineal [23](#)
- modelo de Michaelis Menten
 - en Regresión no lineal [23](#)
- modelo de Morgan-Mercer-Florin
 - en Regresión no lineal [23](#)
- modelo de Peal-Reed
 - en Regresión no lineal [23](#)
- modelo de razón de cuadráticas
 - en Regresión no lineal [23](#)
- modelo de razón de cúbicas
 - en Regresión no lineal [23](#)
- modelo de Richards
 - en Regresión no lineal [23](#)
- modelo de Verhulst
 - en Regresión no lineal [23](#)
- modelo de Von Bertalanffy
 - en Regresión no lineal [23](#)
- modelo de Weibull
 - en Regresión no lineal [23](#)
- modelo log-modificado
 - en Regresión no lineal [23](#)
- modelos de efectos principales
 - en Regresión logística multinomial [10](#)
- Modelos de tiempo de anomalía acelerada paramétrica
 - análisis [35](#)
 - critérios [36](#)
 - Estimación [37](#)
 - exportar [40](#)
 - gráfico [40](#)
 - Imprimir [38](#)
 - modelo [36](#)
 - Predecir [39](#)
- modelos factoriales completos
 - en Regresión logística multinomial [10](#)
- modelos no lineales
 - en Regresión no lineal [23](#)
- modelos personalizados
 - en Regresión logística multinomial [10](#)

P

- potencia relativa de la mediana
 - en el análisis probit [14](#)
- prueba de paralelismo
 - en el análisis probit [14](#)

R

- R cuadrado de McFadden
 - en Regresión logística multinomial [11](#)
- R cuadrado de Nagelkerke
 - en Regresión logística multinomial [11](#)
- R-cuadrado de Cox y Snell
 - en Regresión logística multinomial [11](#)
- razón de verosimilitud

- razón de verosimilitud (*continuación*)
 - bondad de ajuste [11](#)
 - estimación del valor del escalamiento para la dispersión [12](#)
- regresión asintótica
 - en Regresión no lineal [23](#)
- Regresión cuantil
 - critérios [17](#)
 - ejemplo [15](#)
 - estadísticos [15](#)
 - exportación [20](#)
 - guardar [20](#)
 - modelo [18](#)
 - visualización [19](#)
- Regresión de cuadrados mínimos de dos etapas
 - características adicionales del mandato [29](#)
 - covarianzas de los parámetros [29](#)
 - ejemplo [27](#)
 - estadísticos [27](#)
 - guardar variables nuevas [29](#)
 - variables instrumentales [27](#)
- Regresión lineal
 - estimación ponderada [26](#)
 - Regresión de cuadrados mínimos de dos etapas [27](#)
- Regresión lineal Ridge
 - opciones [2](#)
 - partición [2](#)
 - variables [2](#)
- Regresión logística
 - binario [1](#)
 - coeficientes [5](#)
 - contrastes [7](#)
 - covariables categóricas [7](#)
 - covariables de cadena [7](#)
 - definición de la regla de selección [6](#)
 - ejemplo [5](#)
 - establecimiento de reglas [6](#)
 - estadísticos [5](#)
- Regresión Logística
 - características adicionales del mandato [9](#)
 - estadístico de bondad de ajuste de Hosmer-Lemeshow [8](#)
 - estadísticos y gráficos [8](#)
 - guardar variables nuevas [7](#)
 - iteraciones [8](#)
 - medidas de influencia [7](#)
 - métodos de selección de variables [6](#)
 - opciones de visualización [8](#)
 - probabilidad para el método por pasos [8](#)
 - punto de corte para la clasificación [8](#)
 - residuos [7](#)
 - término constante [8](#)
 - valores pronosticados [7](#)
- regresión logística binaria [1, 5](#)
- Regresión logística multinomial
 - características adicionales del mandato [13](#)
 - categoría de referencia [11](#)
 - critérios [12](#)
 - estadísticos [11](#)
 - exportación de información del modelo [13](#)
 - guardar [13](#)
 - Modelos de [10](#)
- regresión logística SPSS [5](#)
- Regresión no lineal

Regresión no lineal (*continuación*)
algoritmo de Levenberg-Marquardt [25](#)
características adicionales del mandato [25](#)
derivadas [24](#)
ejemplo [21](#)
estadísticos [21](#)
estimaciones de programa de arranque [25](#)
función de pérdida [24](#)
guardar variables nuevas [24](#)
interpretación de resultados [25](#)
lógica condicional [22](#)
métodos de estimación [25](#)
modelo segmentado [22](#)
modelos no lineales comunes [23](#)
parámetros [23](#)
programación cuadrática secuencial [25](#)
residuos [24](#)
restricciones para los parámetros [24](#)
valores iniciales [23](#)
valores pronosticados [24](#)

Regresión probit
ejemplo [13](#)
estadísticos [13](#)

regresión restringida
en Regresión no lineal [24](#)

Regresión Ridge del kernel
gráficos [34](#)
guardar [34](#)
parámetros [34](#)
parámetros de cuadrícula [34](#)
pliegues de validación cruzada [34](#)
visualización [34](#)

restricciones para los parámetros
en Regresión no lineal [24](#)

Ridge de kernel
alpha [32](#)
coef0 [32](#)
gamma [32](#)
grado [32](#)
modelo único [32](#)
selección de modelos [32](#)

S

selección hacia delante
en Regresión logística [6](#)

selección por pasos
en Regresión logística [6](#)
en Regresión logística multinomial [10](#)

separación
en Regresión logística multinomial [12](#)

singularidad
en Regresión logística multinomial [12](#)

subdivisión por pasos
en Regresión logística multinomial [12](#)

T

tablas de clasificación
en Regresión logística multinomial [11](#)

tablas de probabilidades de casilla
en Regresión logística multinomial [11](#)

término constante

término constante (*continuación*)
en Regresión lineal [8](#)

V

valor del escalamiento para la dispersión
en Regresión logística multinomial [12](#)

valores de influencia
en Regresión logística [7](#)

