

*IBM SPSS Regression 30*

**IBM**

**Hinweis**

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 51 gelesen werden.

**Produktinformation**

Diese Ausgabe bezieht sich auf Version 30, Release 0, Modifikation 0 von IBM® SPSS Statistics und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuauflage geändert wird.

© **Copyright International Business Machines Corporation .**

---

# Inhaltsverzeichnis

|  |          |
|--|----------|
| <b>Kapitel 1. Regression.....</b>                                      | <b>1</b> |
| Auswählen einer Prozedur für binär logistische Regressionsmodelle..... | 1        |
| Logistische Regression.....  | 2        |
| Logistische Regression: Regel definieren.....                          | 3        |
| Logistische Regression: Methode zur Auswahl von Variablen.....         | 3        |
| Logistische Regression: Kategoriale Variablen definieren.....          | 4        |
| Logistische Regression: Neue Variablen speichern.....                  | 4        |
| Logistische Regression: Optionen.....                                  | 5        |
| Zusätzliche Funktionen beim Befehl LOGISTIC REGRESSION.....            | 6        |
| Multinomiale logistische Regression.....                               | 6        |
| Multinomiale logistische Regression.....                               | 7        |
| Multinomiale logistische Regression: Referenzkategorie.....            | 8        |
| Multinomiale logistische Regression: Statistiken.....                  | 8        |
| Multinomiale logistische Regression: Konvergenzkriterien.....          | 9        |
| Multinomiale logistische Regression: Optionen.....                     | 9        |
| Multinomiale logistische Regression: Speichern.....                    | 10       |
| Zusätzliche Funktionen beim Befehl NOMREG.....                         | 10       |
| Probit-Regression.....   | 10       |
| Probit-Analyse: Bereich definieren.....                                | 12       |
| Probit-Analyse: Optionen.....  | 12       |
| Zusätzliche Funktionen beim Befehl PROBIT.....                         | 12       |
| Quantilregression.....   | 12       |
| Quantilregression: Kriterien.....                                      | 14       |
| Quantilregression: Modell.....   | 15       |
| Quantilregression: Anzeigen.....                                       | 16       |
| Quantilregression: Speichern.....                                      | 17       |
| Quantilregression: Exportieren.....                                    | 18       |
| Nicht lineare Regression.....  | 19       |
| Bedingungslogik (nicht lineare Regression).....                        | 20       |
| Nicht lineare Regression: Parameter.....                               | 20       |
| Nicht lineare Regression: Gängige Modelle.....                         | 21       |
| Nicht lineare Regression: Verlustfunktion.....                         | 21       |
| Nicht lineare Regression: Nebenbedingungen für Parameter.....          | 22       |
| Nicht lineare Regression: Neue Variablen speichern.....                | 22       |
| Nicht lineare Regression: Optionen.....                                | 22       |
| Interpretieren von Ergebnissen der nicht linearen Regression.....      | 23       |
| Zusätzliche Funktionen beim Befehl NLR.....                            | 23       |
| Gewichtungsschätzung.....  | 23       |
| Gewichtungsschätzung: Optionen.....                                    | 25       |
| Zusätzliche Funktionen beim Befehl WLS.....                            | 25       |
| Zweistufige Regressionsmethode der kleinsten Quadrate.....             | 25       |
| Zweistufige kleinste Quadrate: Optionen.....                           | 26       |
| Zusätzliche Funktionen beim Befehl 2SLS.....                           | 26       |
| Codierungsschemata für kategoriale Variablen.....                      | 27       |
| Abweichung.....  | 27       |
| Einfach.....   | 27       |
| Helmert.....   | 28       |
| Differenz.....   | 28       |
| Polynomial.....  | 28       |
| Wiederholt.....  | 29       |
| Spezial.....   | 29       |

|  |           |
|--|-----------|
| Indikator.....   | 30        |
| Kernel-Ridge-Regression.....   | 30        |
| Kernelparameter.....   | 32        |
| Kernel-Ridge-Regression: Optionen.....   | 32        |
| Parametrische beschleunigte Fehlerzeitmodelle.....                                   | 33        |
| Parametrische beschleunigte Fehlerzeitmodelle: Kriterien.....                        | 34        |
| Modelle für parametrische beschleunigte Ausfallzeit: Modell.....                     | 34        |
| Parametrische beschleunigte Fehlerzeitmodelle: Schätzung.....                        | 35        |
| Parametrische beschleunigte Fehlerzeitmodelle: Drucken.....                          | 36        |
| Parametrische Modelle für beschleunigte Ausfallzeiten: Vorhersagen.....              | 37        |
| Modelle für parametrische beschleunigte Ausfallzeit: Diagramm.....                   | 38        |
| Parametrische beschleunigte Fehlerzeitmodelle: Exportieren.....                      | 38        |
| Survival AFT Ereignisse für Statusvariablen definieren.....                          | 39        |
| Parametrische beschleunigte Fehlerzeitmodelle: Kategorie auswählen.....              | 39        |
| Parametrische gemeinsam genutzte Frailty-Modelle.....                                | 40        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Kriterien.....                     | 41        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Modell.....                        | 41        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Schätzung.....                     | 42        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Drucken.....                       | 44        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Vorhersagen.....                   | 44        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Diagramm.....                      | 45        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Exportieren.....                   | 46        |
| Parametrische gemeinsam genutzte Frailty-Modelle: Ereignisse definieren.....         | 46        |
| Parametrische gemeinsam genutzte Frailty-Modelle-Beispiele.....                      | 47        |
| Parametrische gemeinsam genutzte Frailty-Modelle-Fallstudie für Rekurrent-Daten..... | 49        |
| <b>Bemerkungen.....</b>  | <b>51</b> |
| Marken.....  | 52        |
| <b>Index.....</b>  | <b>55</b> |

---

# Kapitel 1. Regression

Die folgenden Regressionsfunktionen sind in SPSS Statistics Standard Edition oder der Option "Regression" enthalten.

## Auswählen einer Prozedur für binär logistische Regressionsmodelle

---

Binär logistische Regressionsmodelle können mit den Prozeduren "Logistische Regression" und "Multinomiale logistische Regression" angepasst werden. Beide Prozeduren bieten Optionen, die in der jeweils anderen nicht vorhanden sind. Eine wichtige theoretische Unterscheidung besteht darin, dass die Prozedur "Logistische Regression" alle Vorhersagen, Residuen, Einflusststatistiken und Tests der Anpassungsgüte anhand von Daten auf der Einzelfallebene erstellt, gleichgültig, auf welche Weise die Daten eingegeben wurden und ob die Anzahl an Kovariatenstrukturen kleiner als die Gesamtzahl der Fälle ist. Die Prozedur "Multinomiale logistische Regression" hingegen aggregiert Fälle intern und bildet Teilgesamtheiten mit identischen Kovariatenstrukturen für die Prädiktoren. Damit werden Vorhersagen, Residuen und Tests zur Anpassungsgüte anhand dieser Teilgesamtheiten erstellt. Wenn alle Prädiktoren kategorial sind oder beliebige stetige Prädiktoren lediglich eine begrenzte Anzahl an Werten annehmen (sodass für jede eindeutige Kovariatenstruktur verschiedene Fälle vorhanden sind), können bei der Methode mit den Teilgesamtheiten gültige Tests der Anpassungsgüte und aussagekräftige Residuen erstellt werden, während dies bei der Methode auf der Grundlage von Einzelfällen nicht möglich wäre.

### Logistische Regression

Sie bietet die folgenden speziellen Funktionen:

- Hosmer-Lemeshow-Anpassungstest für das Modell
- Schrittweise Analyse
- Kontraste für die Definition der Parameterbestimmung des Modells
- Andere Trennwerte für die Klassifikation
- Klassifikationsdiagramme
- Für ein bestimmtes Set von Fällen an ein anderes Set von Fällen angepasstes Modell
- Speichern von Vorhersagen, Residuen und Einflusststatistiken

### Multinomiale logistische Regression

Sie bietet die folgenden speziellen Funktionen:

- Pearson- und Abweichungs-Chi-Quadrat-Anpassungstests für das Modell
- Bestimmung von Teilgesamtheiten zum Gruppieren von Daten für Tests auf Anpassungsgüte
- Auflistung von Häufigkeiten, vorhergesagten Häufigkeiten sowie Residuen nach Teilgesamtheiten
- Korrektur von Varianzschätzungen für die Überstreuung
- Kovarianzmatrix der Parameterschätzungen
- Tests von linearen Parameterkombinationen
- Explizite Festlegung von verschachtelten Modellen
- Anpassen von bedingten logistischen Regressionsmodellen mit 1:1-Übereinstimmung unter Verwendung von differenzierter Variablen

### Hinweise:

- Beide Prozeduren passen ein Modell für Binärdaten an, bei dem es sich um ein verallgemeinertes lineares Modell mit einer Binomialverteilung und einer Logit-Verknüpfungsfunktion handelt. Wenn eine andere Verknüpfungsfunktion für Ihre Daten besser geeignet ist, sollten Sie die Prozedur "Verallgemeinerte lineare Modelle" verwenden.

- Wenn wiederholt Messungen von Binärdaten erforderlich sind oder Datensätze auf andere Weise korreliert werden, sollten Sie die Prozeduren "Verallgemeinerte lineare gemischte Modelle" oder "Verallgemeinerte Schätzungsgleichungen" in Erwägung ziehen.

## Logistische Regression

Die logistische Regression ist für Situationen nützlich, in denen Sie anhand der Werte von Prädiktorvariablen das Vorhandensein oder Nichtvorhandensein einer Eigenschaft oder eines Ergebnisses vorhersagen möchten. Diese Art der Regression verhält sich ähnlich wie ein lineares Regressionsmodell. Sie ist jedoch für Modelle geeignet, in denen die abhängige Variable dichotom ist. Die Koeffizienten der logistischen Regression können verwendet werden, um die Odds-Verhältnisse jeder unabhängigen Variablen im Modell zu schätzen. Die logistische Regression lässt sich auf einen größeren Bereich von Forschungssituationen anwenden als die Diskriminanzanalyse.

**Beispiel.** Welche Verhaltensweisen stellen Risikofaktoren für eine Erkrankung der Herzkranzgefäße dar? Bei einer Stichprobe von Patienten, bei denen Daten hinsichtlich Rauchgewohnheiten, Ernährung, körperliche Betätigung, Alkoholkonsum und Erkrankung der Herzkranzgefäße erhoben wurden, können Sie mit den vier Variablen für die Verhaltensweisen ein Modell erstellen, mit dem das Vorhandensein oder Nichtvorhandensein von Erkrankungen der Herzkranzgefäße in einer Stichprobe von Patienten vorhergesagt werden kann. Das Modell kann dann verwendet werden, um für jeden Faktor Schätzungen der Odds-Verhältnisse abzuleiten, die beispielsweise aussagen, um wie viel wahrscheinlicher Raucher eine Erkrankung der Herzkranzgefäße entwickeln als Nichtraucher.

**Statistiken.** Für jede Analyse: Gesamtzahl der Fälle, ausgewählte Fälle, gültige Fälle. Für jede kategoriale Variable: Parametercodierung. Für jeden Schritt: eingeschlossene und entfernte Variablen, Iterationsverlauf, -2-Log-Likelihood, Güte der Anpassung, Hosmer-Lemeshow-Anpassungstest, Chi-Quadrat für das Modell, Chi-Quadrat für die Verbesserung, Klassifikationstabelle, Diagramm der Korrelationen zwischen Variablen, beobachteten Gruppen und vorhergesagten Wahrscheinlichkeiten, Chi-Quadrat-Residuen. Für jede Variable der Gleichung: Koeffizient ( $B$ ), Standardfehler von  $B$ , Wald-Statistik, geschätztes Odds-Verhältnis ( $\text{Exp}(B)$ ), Konfidenzintervall für  $\text{Exp}(B)$ , Log-Likelihood (falls der Term aus dem Modell entfernt wurde). Für jede nicht in der Gleichung vorhandene Variable: Scorestatistik. Für jeden Fall: beobachtete Gruppe, vorhergesagte Wahrscheinlichkeit, vorhergesagte Gruppenzugehörigkeit, Residuum, standardisiertes Residuum.

**Methoden.** Beim Schätzen des Modells können die Variablen gesammelt im Block oder mit einer der folgenden schrittweisen Methoden eingeschlossen werden: "Vorwärts: Bedingt", "Vorwärts: LQ", "Vorwärts: Wald", "Rückwärts: Bedingt", "Rückwärts: LQ" und "Rückwärts: Wald".

### Erläuterungen der Daten für die logistische Regression

**Daten.** Die abhängige Variable sollte dichotom sein. Unabhängige Variablen können Intervalle oder kategorial sein. Kategoriale Variablen sollten als Dummy oder Indikator codiert sein. (In der Prozedur ist eine Option zum automatischen Umcodieren von kategorialen Variablen verfügbar.)

**Annahmen.** Die logistische Regression beruht nicht so wie die Diskriminanzanalyse auf Annahmen hinsichtlich der Verteilung. Ihre Lösung ist aber möglicherweise stabiler, wenn die Prädiktoren eine multivariate Normalverteilung aufweisen. Wie bei anderen Formen der Regression kann eine Multikollinearität zwischen den Prädiktoren außerdem zu verzerrten Schätzungen und erhöhten Standardfehlern führen. Die Prozedur ist am effektivsten, wenn die Gruppenzugehörigkeit eine echte kategoriale Variable ist. Wenn die Gruppenzugehörigkeit auf den Werten einer stetigen Variablen beruht (z. B. "hoher IQ" bis "niedriger IQ"), empfiehlt sich die lineare Regression, mit der Sie die Vorteile der vielfältigen Informationen von stetigen Variablen nutzen können.

**Verwandte Prozeduren.** Verwenden Sie die Prozedur "Streudiagramm", um Ihre Daten auf Multikollinearität zu sichten. Wenn eine multivariate Normalverteilung und gleiche Varianz-Kovarianz-Matrizen vorliegen, erzielen Sie mit der Prozedur "Diskriminanzanalyse" schneller eine Lösung. Wenn alle Prädiktorvariablen kategorial sind, können Sie auch die Prozedur "Loglinear" verwenden. Wenn die abhängige Variable stetig ist, verwenden Sie die Prozedur "Lineare Regression". Mit der Prozedur "ROC-Kurve" können Sie die Wahrscheinlichkeiten, die mit der Prozedur "Logistische Regression" gespeichert wurden, in einem Diagramm darstellen.

## Berechnen einer logistischen Regressionsanalyse

1. Wählen Sie Folgendes aus den Menüs aus:

**Analysieren > Regression > Binärlogistik...**

2. Wählen Sie eine dichotome abhängige Variable aus. Dies kann eine numerische oder eine Zeichenfolgevariable sein.

3. Wählen Sie mindestens eine Kovariate aus. Um alle Interaktionsterme einzuschließen, wählen Sie erst alle mit der Interaktion verbundenen Variablen und dann **>a\*b>** aus.

Um Variablen in Gruppen (**Blöcken**) einzugeben, wählen Sie erst die Kovariaten für einen Block aus. Klicken Sie dann auf **Weiter**, um einen neuen Block festzulegen. Wiederholen Sie diesen Vorgang, bis alle Blöcke festgelegt wurden.

Sie können außerdem Fälle für die Analyse auswählen. Wählen Sie eine Auswahlvariable aus und geben Sie die Regelkriterien ein.

## Logistische Regression: Regel definieren

Durch die Auswahlregel definierte Fälle werden in die Modellschätzung eingeschlossen. Wenn Sie zum Beispiel eine Variable und **gleich** ausgewählt sowie einen Wert von 5 festgelegt haben, werden nur die Fälle in die Analyse eingeschlossen, für die die ausgewählte Variable den Wert 5 besitzt.

Statistiken und Klassifikationsergebnisse werden sowohl für die ausgewählten als auch für die nicht ausgewählten Fälle erzeugt. Dadurch entsteht ein Mechanismus, mit dem neue Fälle anhand von bereits vorhandenen Daten klassifiziert werden können oder mit dem Sie Ihre Daten in Subsets von Lern- und Testfällen einteilen können, um so eine Gültigkeitsprüfung des generierten Modells durchzuführen.

## Logistische Regression: Methode zur Auswahl von Variablen

Durch die Auswahl der Methode können Sie festlegen, wie unabhängige Variablen in die Analyse eingeschlossen werden. Anhand verschiedener Methoden können Sie eine Vielfalt von Regressionsmodellen mit demselben Set von Variablen erstellen.

- *Eingabe*. Eine Prozedur für die Variablenauswahl, bei der alle Variablen eines Blocks in einem einzigen Schritt aufgenommen werden.
- *Vorwärtsauswahl (Bedingt)*. Eine Methode der schrittweisen Variablenauswahl mit einem Test auf Aufnahme, der auf der Signifikanz der Scorestatistik beruht, und einem Test auf Ausschluss, der auf der Wahrscheinlichkeit einer Likelihood-Quotienten-Statistik beruht, die mit bedingten Parameterschätzungen berechnet wird.
- *Vorwärtsauswahl (Likelihood-Quotient)*. Eine Methode der schrittweisen Variablenauswahl mit einem Test auf Aufnahme, der auf der Signifikanz der Scorestatistik beruht, und einem Test auf Ausschluss, der auf der Wahrscheinlichkeit einer Likelihood-Quotienten-Statistik beruht. Diese basiert hier auf Schätzwerten, die aus dem Maximum einer partiellen Likelihood-Funktion ermittelt werden.
- *Vorwärtsauswahl (Wald)*. Eine Methode der schrittweisen Variablenauswahl mit einem Test auf Aufnahme, der auf der Signifikanz der Scorestatistik beruht, und einem Test auf Ausschluss, der auf der Wahrscheinlichkeit der Wald-Statistik beruht.
- *Rückwärtselimination (Bedingt)*. Schrittweise Rückwärtsauswahl. Der Ausschlusstest basiert auf der Wahrscheinlichkeit der Likelihood-Quotienten-Statistik auf der Grundlage bedingter Parameterschätzungen.
- *Rückwärtselimination (Likelihood-Quotient)*. Schrittweise Rückwärtsauswahl. Der Ausschlusstest basiert auf der Wahrscheinlichkeit der Likelihood-Quotienten-Statistik auf der Grundlage maximaler, partieller Likelihood-Schätzungen.
- *Rückwärtselimination (Wald)*. Schrittweise Rückwärtsauswahl. Der Ausschlusstest basiert auf der Wahrscheinlichkeit der Wald-Statistik.

Die Signifikanzwerte in Ihrer Ausgabe basieren auf der Berechnung eines einzigen Modells. Deshalb sind diese generell ungültig, wenn eine schrittweise Methode verwendet wird.

Alle ausgewählten unabhängigen Variablen werden einem einzigen Regressionsmodell hinzugefügt. Sie können jedoch verschiedene Einschlussmethoden für unterschiedliche Subsets von Variablen angeben. Beispielsweise können Sie einen Block von Variablen durch schrittweises Auswählen und einen zweiten Block durch Vorwärtsselektion in das Regressionsmodell einschließen. Um einem Regressionsmodell einen zweiten Block von Variablen hinzuzufügen, klicken Sie auf **Weiter**.

## Logistische Regression: Kategoriale Variablen definieren

Sie können festlegen, wie die Prozedur "Logistische Regression" kategoriale Variablen verarbeitet:

**Kovariaten.** Enthält eine Liste aller im Hauptdialogfeld bestimmten Kovariaten, alleine oder als Teil einer Interaktion in einer beliebigen Schicht. Falls einige der Variablen Zeichenfolgevariablen oder kategoriale Variablen sind, können Sie diese nur als kategoriale Kovariaten verwenden.

**Kategoriale Kovariaten.** Enthält die Variablen, die als kategorial festgelegt wurden. Bei jeder Variablen wird in Klammern die zu verwendende Kontrastcodierung angezeigt. Zeichenfolgevariablen (durch das Symbol < nach dem Namen gekennzeichnet) wurden bereits in die Liste "Kategoriale Kovariaten" übernommen. Wählen Sie gegebenenfalls andere kategoriale Kovariaten aus der Liste "Kovariaten" aus und verschieben Sie diese in die Liste "Kategoriale Kovariaten".

**Kontrast ändern.** Hiermit können Sie die Kontrastmethode ändern. Die folgenden Kontrastmethoden sind verfügbar:

- **Indikator.** Die Kontraste kennzeichnen das Vorhandensein oder Nichtvorhandensein einer Kategoriezugehörigkeit. Die Referenzkategorie wird in der Kontrastmatrix als Zeile mit Nullen dargestellt.
- **Einfach.** Jede Kategorie der Vorhersagevariablen mit Ausnahme der Referenzkategorie wird mit der Referenzkategorie verglichen.
- **Differenz.** Jede Kategorie der Prädiktorvariablen mit Ausnahme der ersten Kategorie wird mit dem durchschnittlichen Effekt der vorigen Kategorien verglichen. Dies ist auch als umgekehrte Helmert-Kontraste bekannt.
- **Helmert.** Jede Kategorie der Prädiktorvariablen mit Ausnahme der letzten Kategorie wird mit dem durchschnittlichen Effekt der nachfolgenden Kategorien verglichen.
- **Wiederholt.** Jede Kategorie der Vorhersagevariablen mit Ausnahme der letzten Kategorie wird mit der nächsten Kategorie verglichen.
- **Polynomial.** Orthogonale polynomiale Kontraste. Es wird angenommen, dass zwischen den Kategorien die gleichen Abstände vorliegen. Polynomiale Kontraste sind nur für numerische Variablen verfügbar.
- **Abweichung.** Jede Kategorie der Prädiktorvariablen mit Ausnahme der Referenzkategorie wird mit dem Gesamteffekt verglichen.

Wenn Sie **Abweichung**, **Einfach** oder **Indikator** ausgewählt haben, wählen Sie entweder **Erste** oder **Letzte** als Referenzkategorie aus. Beachten Sie, dass die Methode erst geändert wird, wenn Sie auf **Ändern** klicken.

Die Zeichenfolgekovariaten müssen kategoriale Kovariaten sein. Um eine Zeichenfolgevariable aus der Liste "Kategoriale Kovariaten" zu entfernen, müssen Sie alle Terme, welche diese Variable enthalten, aus der Liste "Kovariaten" im Hauptdialogfeld entfernen.

## Logistische Regression: Neue Variablen speichern

Sie können die Ergebnisse der logistischen Regression als neue Variablen im aktiven Dataset speichern.

**Vorhergesagte Werte.** Speichert vom Modell vorhergesagte Werte. Die verfügbaren Optionen sind "Wahrscheinlichkeiten" und "Gruppenzugehörigkeit".

- **Wahrscheinlichkeiten** Speichert für jeden Fall die vorhergesagte Wahrscheinlichkeit des Auftretens des Ereignisses. In der Ausgabe zeigt eine Tabelle die Namen und Inhalte aller neuen Variablen an. Das "Ereignis" ist die Kategorie der abhängigen Variablen mit dem höheren Wert. Wenn die abhängige Variable beispielsweise die Werte 0 und 1 annimmt, wird die vorhergesagte Wahrscheinlichkeit der Kategorie 1 gespeichert.



- *Vorhergesagte Gruppenzugehörigkeit.* Die Gruppe mit der größten A-posteriori-Wahrscheinlichkeit basierend auf Diskriminanzscores. Die Gruppe, der der Fall durch das Modell zugeordnet wird.

**Einfluss.** Speichert Werte aus Statistiken, welche den Einfluss der Fälle auf die vorhergesagten Werte messen. Die Optionen "Cook", "Hebelwerte" und "DfBeta" sind verfügbar.

- *Cook.* Die logistische Regression analog zur Einflusstatistik von Cook. Ein Maß dafür, wie stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen würde.
- *Hebelwert.* Der relative Einfluss einer jeden Beobachtung auf die Anpassungsgüte eines Modells.
- *DfBetas.* Die Differenz im Betawert ist die Änderung des Regressionskoeffizienten, die sich aus dem Ausschluss eines bestimmten Falls ergibt. Für jeden Term im Modell, einschließlich der Konstanten, wird ein Wert berechnet.

**Residuen.** Speichert die Residuen. Die Optionen "Nicht standardisiert", "Logit", "Studentisiert", "Standardisiert" und "Abweichung" sind verfügbar.

- *Nicht standardisierte Residuen.* Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- *Logit-Residuum.* Das Residuum für den Fall, wenn er in der Logit-Skala vorhergesagt wird. Das Logit-Residuum ist das Residuum dividiert durch die vorhergesagte Wahrscheinlichkeit mal 1 minus der vorhergesagten Wahrscheinlichkeit.
- *Studentisiertes Residuum.* Die Änderung in der Modellabweichung bei Ausschluss eines Falles.
- *Standardisierte Residuen.* Das Residuum dividiert durch eine Schätzung der Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- *Deviance.* Residuen basierend auf der Modellabweichung.

**Modellinformationen in XML-Datei exportieren.** Parameterschätzungen und (wahlweise) ihre Kovarianzen werden in die angegebene Datei exportiert. Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden. .

## Logistische Regression: Optionen

Sie können Optionen für die logistische Regressionsanalyse festlegen.

**Statistiken und Diagramme.** Mit dieser Funktion können Sie Statistiken und Diagramme anfordern. Die Optionen "Klassifikationsdiagramme", "Hosmer-Lemeshow-Anpassungsstatistik", "Fallweise Auflistung der Residuen", "Korrelationen der Schätzungen", "Iterationsverlauf" und "Konfidenzint. für  $\text{Exp}(B)$ " sind verfügbar. Wählen Sie im Gruppenfeld "Anzeigen" eine der Optionen aus, sodass Statistiken und Diagramme entweder "Bei jedem Schritt" oder nur das endgültige Modell "Beim letzten Schritt" angezeigt wird.

- *Hosmer-Lemeshow-Anpassungsstatistik.* Diese Statistik zur Anpassungsgüte ist robuster als die traditionelle Statistik zur Anpassungsgüte, die in der logistischen Regression verwendet wird, insbesondere für Modelle mit stetigen Kovariaten und Studien mit kleinen Stichprobenumfängen. Sie basiert auf der Gruppierung von Fällen in Risikodezile und dem Vergleich der beobachteten Wahrscheinlichkeit mit der erwarteten Wahrscheinlichkeit innerhalb der einzelnen Dezile.

**Wahrscheinlichkeit für schrittweise Methode.** Mit dieser Funktion können Sie die Kriterien festlegen, nach denen die Variablen in die Gleichung aufgenommen oder aus dieser entfernt werden. Sie können die Kriterien zur Aufnahme oder zum Entfernen der Variablen festlegen.

- *Wahrscheinlichkeit für schrittweise Methode.* Eine Variable wird in das Modell aufgenommen, wenn die Wahrscheinlichkeit ihrer Scorestatistik kleiner als der Eingabewert ist. Sie wird entfernt, wenn die Wahrscheinlichkeit größer als der Ausschlusswert ist. Um die Standardeinstellungen zu überschreiben, geben Sie positive Werte für Aufnahme- und Ausschlusswert ein. Der Aufnahmewert muss kleiner sein als der Ausschlusswert.

**Klassifikationstrennwert.** Hiermit können Sie den Trennwert für die Klassifikation von Fällen festlegen. Fälle mit vorhergesagten Werten, die den Klassifikationstrennwert übersteigen, werden als positiv, vor-

hergesagte Werte, die unter dem Schwellenwert liegen, als negativ klassifiziert. Um die Standardeinstellung zu ändern, geben Sie einen Wert zwischen 0,01 und 0,99 ein.

**Maximalzahl der Iterationen.** Hiermit können Sie festlegen, wie oft das Modell vor Abschluss iteriert.

**Konstante in Modell einschließen.** Mit dieser Funktion können Sie angeben, ob das Modell einen konstanten Term enthalten soll. Wenn diese Option inaktiviert ist, ist der konstante Term gleich 0.

## Zusätzliche Funktionen beim Befehl LOGISTIC REGRESSION

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Die fallweise Ausgabe kann mit den Werten oder Beschriftungen einer Variablen beschriftet werden.
- Der Abstand zwischen Objekten in den Iterationsberichten kann angegeben werden. Statt einer Ausgabe der Parameterschätzungen nach jeder Iteration können Parameterschätzungen nach jeder  $n$ -ten Iteration angefordert werden.
- Die Kriterien zum Beenden der Iteration können geändert und die Redundanz geprüft werden.
- Für fallweise Listen kann eine Variablenliste festgelegt werden.
- Bei der Verarbeitung können die Daten jeder Gruppe einer aufgeteilten Datei in einer externen temporären Datei zwischengespeichert werden. Dadurch wird Speicherplatz gespart.

Vollständige Syntaxinformationen finden Sie unter *Befehlssyntaxreferenz*.

## Multinomiale logistische Regression

---

Multinomiale logistische Regression bietet sich in Situationen an, in denen Sie Subjekte anhand von Variablen aus einem Set von Prädiktorvariablen klassifizieren möchten. Diese Art von Regression gleicht einer logistischen Regression, ist jedoch allgemeiner, da die abhängige Variable nicht auf zwei Kategorien beschränkt ist.

**Beispiel.** Filmstudios sind an Vorhersagen zu der Art von Filmen interessiert, die sich Kinogänger am wahrscheinlichsten ansehen, damit Filme besser vermarktet werden können. Mit einer multinomialen logistischen Regression kann bestimmt werden, zu welchem Grad das Alter, das Geschlecht und die Beziehungssituation einer Person den bevorzugten Filmtyp beeinflusst. Studios können die Werbekampagne für einen bestimmten Film dann auf die Zielgruppen ausrichten, die den Film wahrscheinlich sehen werden.

**Statistiken.** Iterationsverlauf, Parameterkoeffizienten, asymptotische Kovarianz- und Korrelationsmatrizen, Likelihood-Quotienten-Tests für Modelleffekte und partielle Effekte, -2-Log-Likelihood, Pearson- und Abweichungs-Chi-Quadrat-Anpassungstests,  $R^2$  nach Cox und Snell, Nagelkerke und McFadden. Klassifikation: beobachtete Häufigkeiten gegenüber vorhergesagten Häufigkeiten nach Antwortkategorie. Kreuztabelle: beobachtete Häufigkeiten und vorhergesagte Häufigkeiten (mit Residuen) und Anteile nach Kovariatenstruktur und Antwortkategorie.

**Methoden.** Ein multinomiales Logit-Modell wird für das gesättigte Modell oder ein benutzerdefiniertes Modell angepasst. Die Parameterschätzung wird durch einen iterativen Maximum-Likelihood-Algorithmus durchgeführt.

### Erläuterungen der Daten für die multinomiale logistische Regression

**Daten.** Die abhängige Variable sollte kategorial sein. Unabhängige Variablen können Faktoren oder Kovariaten sein. Im Allgemeinen sollten Faktoren kategoriale Variablen und Kovariaten stetige Variablen sein.

**Annahmen.** Es wird angenommen, dass das Odds-Verhältnis von zwei beliebigen Kategorien unabhängig von allen anderen Antwortkategorien ist. Wenn beispielsweise ein neues Produkt auf dem Markt eingeführt wird, besagt diese Annahme, dass die Marktanteile aller anderen Produkte proportional gleichermaßen verändert werden. Zusätzlich wird angenommen, dass die Antworten bei gegebener Kovariatenstruktur unabhängige multinomiale Variablen darstellen.

## Berechnen einer multinomialen logistischen Regression

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Regression > Multinomiale logistische Regression ...**

2. Wählen Sie eine abhängige Variable aus.

3. Faktoren sind optional und können sowohl numerisch als auch kategorial sein.

4. Kovariaten sind optional, müssen jedoch, falls festgelegt, numerisch sein.

## Multinomiale logistische Regression

Die Prozedur "Multinomiale logistische Regression" führt standardmäßig zu einem Modell mit den Haupteffekten der Faktoren und Kovariaten, Sie können jedoch in diesem Dialogfeld auch ein benutzerdefiniertes Modell angeben oder schrittweise Modellauswahl anfordern.

**Modell angeben.** Ein Modell mit Haupteffekten enthält die Haupteffekte der Faktoren und Kovariaten, aber keine Interaktionseffekte. Ein gesättigtes Modell enthält alle Haupteffekte sowie alle Interaktionen zwischen den Faktoren. Es enthält keine Kovariateninteraktionen. Sie können ein benutzerdefiniertes Modell erstellen, um Subsets von Interaktionen zwischen Faktoren oder Kovariaten zu bestimmen oder die schrittweise Auswahl der Modellterme anzufordern.

**Faktoren & Kovariaten.** Die Faktoren und Kovariaten werden aufgelistet.

**Terme für erzwungenen Einschluss.** Die in die Liste für den erzwungenen Einschluss aufgenommenen Terme sind immer im Modell enthalten.

**Terme für schrittweisen Einschluss.** Die zu der Liste für den schrittweisen Einschluss hinzugefügten Terme werden gemäß einer der folgenden vom Benutzer auszuwählenden schrittweisen Methoden in das Modell aufgenommen:

- **Vorwärtseingabe.** Bei dieser Methode sind zunächst keine Terme für schrittweisen Einschluss im Modell enthalten. Bei jedem Schritt wird der wichtigste Term zum Modell hinzugefügt, bis keiner der Terme für schrittweisen Einschluss, die nicht im Modell enthalten sind, einen statistisch signifikanten Beitrag leisten würde, wenn er in das Modell aufgenommen würde.
- **Rückwärtselimination.** Bei dieser Methode werden zunächst alle in der Liste für die schrittweise Auswahl angegebenen Terme in das Modell eingegeben. Bei jedem Schritt wird der am wenigsten signifikante Term für schrittweisen Einschluss aus dem Modell entfernt, bis alle verbleibenden Terme für schrittweisen Einschluss einen statistisch signifikanten Beitrag zum Modell leisten.
- **Schrittweise vorwärts** Diese Methode beginnt mit dem Modell, das mit der Methode "Vorwärtseingabe" ausgewählt würde. Dann wechselt der Algorithmus zwischen Rückwärtselimination bei den Termen für schrittweisen Einschluss im Modell und Vorwärtseingabe bei den nicht in das Modell aufgenommenen Termen. Dieser Vorgang wird so lange fortgesetzt, bis keine Terme mehr die Kriterien für Eingabe oder Ausschluss erfüllen.
- **Rückwärtsgerichtet schrittweise.** Diese Methode beginnt mit dem Modell, das mit der Methode "Rückwärtselimination" ausgewählt würde. Dann wechselt der Algorithmus zwischen Vorwärtseingabe bei den nicht in das Modell aufgenommenen Termen und Rückwärtselimination bei den Terme für schrittweisen Einschluss im Modell. Dieser Vorgang wird so lange fortgesetzt, bis keine Terme mehr die Kriterien für Eingabe oder Ausschluss erfüllen.

**Konstanten Term in Modell einschließen.** Hiermit können Sie einen konstanten Term für das Modell ein- oder ausschließen.

## Bedingungen erstellen

Für die ausgewählten Faktoren und Kovariaten:

**Interaktion.** Hiermit wird der Interaktionsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt.

**Haupteffekte.** Legt einen Haupteffektterm für jede ausgewählte Variable an.

**Alle 2-Wege.** Hiermit werden alle möglichen Zweiwegeinteraktionen der ausgewählten Variablen erzeugt.

**Alle 3-Wege.** Hiermit werden alle möglichen Dreiwegeinteraktionen der ausgewählten Variablen erzeugt.

**Alle 4-Wege.** Hiermit werden alle möglichen Vierwegeinteraktionen der ausgewählten Variablen erzeugt.

**Alle 5-Wege.** Hiermit werden alle möglichen Fünfwegeinteraktionen der ausgewählten Variablen erzeugt.

## Multinomiale logistische Regression: Referenzkategorie

Standardmäßig wird bei der Prozedur "Multinomiale logistische Regression" die letzte Kategorie als Referenzkategorie verwendet. In diesem Dialogfeld können Sie die Referenzkategorie und die Reihenfolge der Kategorien festlegen.

**Referenzkategorie.** Geben Sie die erste, letzte oder eine benutzerdefinierte Kategorie an.

**Kategoriereihenfolge.** In aufsteigender Reihenfolge; der niedrigste Wert steht für die erste Kategorie und der höchste Wert für die letzte Kategorie. In absteigender Reihenfolge; der höchste Wert steht für die erste Kategorie und der niedrigste Wert für die letzte Kategorie.

## Multinomiale logistische Regression: Statistiken

Sie können die folgenden Statistiken für die multinomiale logistische Regression festlegen:

**Zusammenfassung der Fallverarbeitung.** Diese Tabelle enthält Informationen zu den angegebenen kategorialen Variablen.

**Modell.** Statistiken für das Gesamtmodell.

- **Pseudo-R-Quadrat.** Gibt  $R^2$  nach Cox und Snell, Nagelkerke und McFadden aus.
- **Zusammenfassung der Schritte.** Diese Tabelle bietet einen Überblick über die Effekte, die bei einer schrittweisen Methode in den einzelnen Schritten eingegeben bzw. entfernt wurden. Diese Tabelle wird nur dann angezeigt, wenn im Dialogfeld Modell ein schrittweises Modell angegeben wurde.
- **Information zur Modellanpassung.** In dieser Tabelle werden die angepassten Modelle mit den Modellen mit ausschließlich konstanten Termen und den Nullmodellen verglichen.
- **Informationskriterien.** Diese Tabelle gibt das Akaike-Informationskriterium (AIC) und das Bayes-Informationskriterium von Schwarz (BIC) aus.
- **Zellwahrscheinlichkeiten.** Gibt eine Tabelle der beobachteten und erwarteten Häufigkeiten (mit Residuen) und Anteile nach Kovariatenstruktur und Antwortkategorie aus.
- **Klassifikationstabelle.** Gibt eine Tabelle der beobachteten gegenüber den vorhergesagten Antworten aus.
- **Chi-Quadrat-Statistik für Anpassungsgüte.** Gibt die Chi-Quadrat-Statistik nach Pearson und die Likelihood-Quotienten-Chi-Quadrat-Statistik aus. Die Statistiken werden für die Kovariatenstrukturen berechnet, die durch alle Faktoren und Kovariaten oder ein benutzerdefiniertes Subset der Faktoren und Kovariaten definiert sind.
- **Monotonizitätsmaße.** Zeigt eine Tabelle mit Informationen zur Anzahl der konkordanten Paare, der diskordanten Paare und der gebundenen Paare an. In dieser Tabelle werden außerdem die Werte Somers-D, Goodman-und-Kruskal-Gamma, Kendall-Tau-a und Konkordanzindex C angezeigt.

**Parameter.** Statistiken in Bezug auf die Modellparameter.

- **Schätzungen.** Gibt Schätzungen der Modellparameter mit einem benutzerdefinierten Konfidenzniveau aus.
- **Likelihood-Quotienten-Test.** Gibt Likelihood-Quotienten-Tests für die partiellen Effekte des Modells aus. Der Test für das gesamte Modell wird automatisch ausgegeben.
- **Asymptotische Korrelationen.** Gibt eine Matrix der Parameterschätzungskorrelationen aus.
- **Asymptotische Kovarianz.** Gibt eine Matrix der Parameterschätzungskovarianzen aus.

**Teilgesamtheiten definieren.** Hiermit können Sie ein Subset aus den Faktoren und Kovariaten wählen und so die Kovariatenstrukturen definieren, die von allen Zellenwahrscheinlichkeiten und den Tests zur Anpassungsgüte verwendet werden.

## Multinomiale logistische Regression: Konvergenzkriterien

Sie können die folgenden Kriterien für die multinomiale logistische Regression angeben:

**Iterationen.** Hier können Sie angeben, wie oft der Algorithmus höchstens durchlaufen werden soll. Außerdem können Sie die Maximalzahl der Schritte in der Schritthalbierung und die Konvergenztoleranzen für Änderungen an Log-Likelihood und Parametern festlegen und bestimmen, wie oft der Iterationsverlauf ausgegeben werden soll. Ferner können Sie festlegen, wie oft die Daten auf vollständige oder quasi vollständige Trennung geprüft werden sollen.

- **Log-Likelihood-Konvergenz.** Es wird angenommen, dass eine Konvergenz vorliegt, wenn die relative Änderung in der Log-Likelihood-Funktion kleiner als der angegebene Wert ist. Wenn der Wert gleich 0 ist, wird dieses Kriterium nicht angewendet. Geben Sie einen nicht negativen Wert an.
- **Parameterkonvergenz.** Es wird angenommen, dass eine Konvergenz vorliegt, wenn die absolute Änderung in den Parameterschätzungen kleiner als dieser Wert ist. Wenn der Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

**Delta.** Ermöglicht es Ihnen, einen nicht negativen Wert kleiner als 1 anzugeben. Dieser Wert wird zu jeder leeren Zelle der Kreuztabelle der Antwortkategorie durch Kovariatenstruktur hinzugefügt. Damit kann der Algorithmus stabilisiert und eine Verzerrung der Schätzungen verhindert werden.

**Toleranz für Prüfung auf Singularität.** Hier können Sie die Toleranz angeben, die bei der Prüfung auf Singularität verwendet wird.

## Multinomiale logistische Regression: Optionen

Sie können die folgenden Optionen für die multinomiale logistische Regression festlegen:

**Streuungsmaßstab.** Hiermit können Sie den Skalierungswert für die Streuung angeben, mit dem die Schätzung der Parameterkovarianzmatrix korrigiert wird. Bei der Option **Abweichung** wird der Skalierungswert unter Verwendung der Abweichungsfunktion (Likelihood-Quotienten-Chi-Quadrat) geschätzt. Bei der Option **Pearson** wird der Skalierungswert unter Verwendung der Chi-Quadrat-Statistik nach Pearson geschätzt. Sie können auch einen Skalierungswert angeben. Hierbei muss es sich um einen positiven numerischen Wert handeln.

**Schrittweise Optionen.** Mit diesen Optionen können Sie die statistischen Kriterien festlegen, wenn schrittweise Methoden zur Erstellung eines Modells verwendet werden. Diese werden nur dann berücksichtigt, wenn im Dialogfeld Modell ein schrittweises Modell angegeben wurde.

- **Wahrscheinlichkeit für Einschluss.** Hierbei handelt es sich um die Wahrscheinlichkeit der Likelihood-Quotienten-Statistik für die Variablenaufnahme. Je größer die angegebene Wahrscheinlichkeit, desto leichter wird eine Variable in das Modell aufgenommen. Dieses Kriterium wird nur dann berücksichtigt, wenn die Methode "Vorwärtseingabe" "Vorwärtsgerichtet schrittweise" oder "Rückwärtsgerichtet schrittweise" ausgewählt ist.
- **Einschlusstest.** Hierbei handelt es sich um das Verfahren zur Aufnahme von Terms bei schrittweisen Methoden. Wählen Sie zwischen Likelihood-Quotienten-Test und Scoretest. Dieses Kriterium wird nur dann berücksichtigt, wenn die Methode "Vorwärtseingabe" "Vorwärtsgerichtet schrittweise" oder "Rückwärtsgerichtet schrittweise" ausgewählt ist.
- **Wahrscheinlichkeit für Ausschluss.** Hierbei handelt es sich um die Wahrscheinlichkeit der Likelihood-Quotienten-Statistik für den Variablenausschluss. Je größer die angegebene Wahrscheinlichkeit, desto leichter verbleibt eine Variable im Modell. Dieses Kriterium wird nur dann berücksichtigt, wenn die Methode "Rückwärtselimination", "Vorwärtsgerichtet schrittweise" oder "Rückwärtsgerichtet schrittweise" ausgewählt ist.
- **Ausschlusstest.** Hierbei handelt es sich um das Verfahren zur Entfernung von Terms bei schrittweisen Methoden. Wählen Sie zwischen Likelihood-Quotienten-Test und Wald-Test. Dieses Kriterium wird nur

dann berücksichtigt, wenn die Methode "Rückwärtselimination", "Vorwärtsgerichtet schrittweise" oder "Rückwärtsgerichtet schrittweise" ausgewählt ist.

- **Kleinste Anzahl schrittweise ermittelter Effekte im Modell.** Bei Verwendung der Methode "Rückwärtselimination" oder "Rückwärtsgerichtet schrittweise" wird hier die Mindestzahl der Terme angegeben, die in das Modell aufgenommen werden. Der konstante Term wird nicht als Modellterm gewertet.
- **Größte Anzahl schrittweise ermittelter Effekte im Modell.** Bei Verwendung der Methode "Vorwärtsseingabe" oder "Vorwärtsgerichtet schrittweise" wird hier die Mindestzahl der Terme angegeben, die in das Modell aufgenommen werden. Der konstante Term wird nicht als Modellterm gewertet.
- **Eingabe und Entfernung von Termen hierarchisch einschränken.** Mit dieser Option können Sie auswählen, ob Beschränkungen für die Aufnahme von Modelltermen gelten sollen. Bei der Hierarchieoption müssen für jeden aufzunehmenden Term alle Terme niedrigerer Ordnung, die zum aufzunehmenden Term gehören, zuvor in das Modell aufgenommen werden. Beispiel: Wenn die Hierarchievoraussetzung in Kraft ist, müssen die Faktoren *Familienstand* und *Geschlecht* im Modell enthalten sein, bevor der Interaktionsterm *Familienstand\*Geschlecht* hinzugefügt werden kann. Mit den drei Optionsfeldoptionen wird die Rolle der Kovariaten bei der Ermittlung der Hierarchie festgelegt.

## Multinomiale logistische Regression: Speichern

Mit dem Dialogfeld "Speichern" können Sie Variablen in der Arbeitsdatei speichern und Modellinformationen in eine externe Datei exportieren.

**Gespeicherte Variablen.** Die folgenden Variablen können gespeichert werden:

- **Geschätzte Antwortwahrscheinlichkeiten.** Hierbei handelt es sich um die geschätzten Wahrscheinlichkeiten, dass eine Faktor-/Kovariaten-Struktur in eine Antwortkategorie klassifiziert wird. Es gibt ebenso viele geschätzte Wahrscheinlichkeiten wie Kategorien der Antwortvariablen. Bis zu 25 davon können gespeichert werden.
- **Vorhergesagte Kategorie.** Hierbei handelt es sich um die Antwortkategorie mit der größten erwarteten Wahrscheinlichkeit für eine Faktor-/Kovariaten-Struktur.
- **Vorhergesagte Kategoriewahrscheinlichkeiten.** Hierbei handelt es sich um das Maximum an geschätzten Antwortwahrscheinlichkeiten.
- **Tatsächliche Kategoriewahrscheinlichkeit.** Hierbei handelt es sich um die geschätzte Wahrscheinlichkeit, dass eine Faktor-/Kovariaten-Struktur in die beobachtete Kategorie klassifiziert wird.

**Modellinformationen in XML-Datei exportieren.** Parameterschätzungen und (wahlweise) ihre Kovarianzen werden in die angegebene Datei exportiert. Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden. .

## Zusätzliche Funktionen beim Befehl NOMREG

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Sie können die Referenzkategorie der abhängigen Variablen angeben.
- Sie können benutzerdefiniert fehlende Werte einschließen.
- Sie können Hypothesentests durch Festlegen von Nullhypothesen als lineare Parameterkombinationen anpassen.

Vollständige Syntaxinformationen finden Sie im Handbuch *Befehlssyntaxreferenz*.

## Probit-Regression

---

Diese Prozedur misst die Beziehung zwischen der Stärke eines Stimulus und dem Anteil der Fälle, die eine bestimmte Antwort auf den Stimulus zeigen. Sie ist nützlich in Situationen mit dichotomer Ausgabe, die wahrscheinlich von den Faktorstufen einiger unabhängiger Variablen beeinflusst oder verursacht wird, und passt insbesondere zu experimentellem Datenmaterial. Diese Prozedur ermöglicht Ihnen, die Stärke eines Stimulus zu schätzen, der notwendig ist, um einen bestimmten Anteil an Responses zu erzielen, beispielsweise die mittlere effektive Dosis.

**Beispiel.** Wie wirksam ist ein neues Schädlingsbekämpfungsmittel gegen Ameisen und wie hoch sollte die Konzentration des Mittels sinnvollerweise sein? Sie können ein Experiment durchführen, bei dem Stichproben von Ameisen verschiedenen Konzentrationen des Schädlingsbekämpfungsmittels ausgesetzt sind. Zeichnen Sie dabei die Anzahl der vernichteten Ameisen und die Anzahl der dem Mittel ausgesetzten Ameisen auf. Wenden Sie anschließend die Probit-Regression auf diese Daten an. Sie können damit die Stärke der Beziehung zwischen der Konzentration und der Anzahl vernichteter Ameisen ermitteln. Sie können außerdem die geeignete Konzentration des Bekämpfungsmittels bestimmen, um mit Sicherheit beispielsweise 95 % der dem Mittel ausgesetzten Ameisen zu vernichten.

**Statistik** Regressionskoeffizienten und Standardfehler, konstanter Term und Standardfehler, Chi-Quadrat-Anpassungstest nach Pearson, beobachtete und erwartete Häufigkeiten sowie Konfidenzintervalle für effektive Faktorstufen der unabhängigen Variablen. Diagramme: transformierte Antwortdiagramme.

Diese Prozedur verwendet die Algorithmen, die in NPSOL<sup>®</sup> von Gill, Murray, Saunders & Wright vorgeschlagen und implementiert wurden, um die Modellparameter zu schätzen.

## Erläuterungen der Daten für die Probit-Regression

**Daten.** Für jeden Wert der unabhängigen Variablen oder für jede Wertekombination mehrerer unabhängiger Variablen sollte Ihre Antwortvariable die Häufigkeit der Fälle mit den Werten sein, welche von Interesse sind. Die Variable mit der beobachteten Gesamtzahl sollte die Gesamtzahl der Fälle mit diesen Werten für die unabhängige Variable sein. Die Faktorvariable sollte kategorial und als Ganzzahl codiert sein.

**Annahmen.** Die Beobachtungen müssen unabhängig sein. Wenn Sie im Verhältnis zur Anzahl der Beobachtungen über eine große Anzahl von Werten für die unabhängige Variablen verfügen, wie dies bei empirischen Studien der Fall sein kann, ist es möglich, dass die Chi-Quadrat- und die Anpassungsstatistik nicht gültig sind.

**Verwandte Prozeduren.** Die Probit-Analyse ist eng verwandt mit der logistischen Regression. Wenn Sie die Logit-Transformation auswählen, berechnet diese Prozedur im wesentlichen eine logistische Regression. Die Probit-Analyse ist in der Regel für geplante Experimente geeignet. Die logistische Regression ist dagegen eher für empirische Studien geeignet. Die Unterschiede in der Ausgabe bringen diese unterschiedlichen Betonungen zum Ausdruck. Die Prozedur "Probit-Analyse" ermittelt Schätzungen der effektiven Werte für verschiedene Rücklaufquoten (einschließlich der mittleren effektiven Dosis), während die Prozedur "Logistische Regression" Schätzungen der Odds-Verhältnisse für unabhängige Variablen berechnet.

## Berechnen einer Probit-Regressionsanalyse

1. Wählen Sie in den Menüs Folgendes aus:

**Regression > Probit > analysieren...**

2. Wählen Sie eine Variable für die Antworthäufigkeit. Diese Variable zeigt die Anzahl der Fälle an, die auf den Teststimulus reagieren. Die Werte dieser Variablen dürfen nicht negativ sein.

3. Wählen Sie eine Variable für die beobachtete Gesamtzahl aus. Diese Variable zeigt die Anzahl der Fälle an, auf welche der Stimulus angewendet wurde. Die Werte dieser Variablen dürfen nicht negativ und außerdem für jeden Fall nicht kleiner als die Werte der Variablen für die Antworthäufigkeit sein.

Wahlweise können Sie eine Faktorvariable auswählen. Wenn Sie dies tun, verwenden Sie **Bereich definieren**, um den Bereich der Gruppen zu definieren.

4. Wählen Sie mindestens eine Kovariate aus. Diese Variable enthält die Faktorstufe des Stimulus, der in jeder Beobachtung verwendet wird. Wenn Sie die Kovariate transformieren möchten, wählen Sie eine Transformation aus dem Dropdown-Listefeld **Transformation** aus. Falls keine Transformation angewendet wird und eine Kontrollgruppe vorhanden ist, wird die Kontrollgruppe in die Analyse einbezogen.

5. Wählen Sie das Modell **Probit** oder **Logit** aus.

### **Probit-Modell**

Wendet die Probit-Transformation (die Inverse der kumulativen Standardnormalverteilungsfunktion) auf die Anteile der Antworten an.

### **Logit-Modell**

Wendet die Logit-Transformation (Log-Odds-Transformation) auf die Antwortverhältnisse an.

## **Probit-Analyse: Bereich definieren**

Hiermit können Sie die Stufen der Faktorvariablen festlegen, die analysiert werden soll. Die Faktorstufen müssen als aufeinander folgende ganze Zahlen codiert sein. Alle Stufen in dem von Ihnen festgelegten Bereich werden analysiert.

## **Probit-Analyse: Optionen**

Sie können Optionen für die Probit-Analyse festlegen.

**Statistik.** Die folgenden optionalen Statistiken stehen zur Verfügung: Häufigkeiten, mittlere effektive Dosis, Parallelitätstest und Fiduzialkonfidenzintervalle.

- *Mittlere effektive Dosis.* Zeigt das Verhältnis der mittleren Potenzen für jedes Paar von Faktorstufen an. Zeigt außerdem 95-%-Konfidenzgrenzen für jede mittlere effektive Dosis an. Mittlere effektive Dosen stehen nicht zur Verfügung, wenn keine Faktorvariable oder mehr als eine Kovariate vorhanden ist.
- *Parallelitätstest.* Ein Test der Hypothese, dass alle Faktorstufen die gleiche Steigung aufweisen.
- *Fiduzial-Konfidenzintervalle.* Konfidenzintervalle für die Dosierung eines Wirkstoffs, die erforderlich ist, um eine bestimmte Wahrscheinlichkeit für die Wirkung zu erhalten.

Wenn Sie mehr als eine Kovariate ausgewählt haben, sind die Fiduzialkonfidenzintervalle und die mittlere effektive Dosis nicht verfügbar. Die mittlere effektive Dosis und der Parallelitätstest sind nur verfügbar, wenn Sie eine Faktorvariable ausgewählt haben.

**Natürliche Rücklaufquote.** Hiermit können Sie eine natürliche Rücklaufquote angeben, selbst wenn der Stimulus fehlt. Verfügbare Alternativen sind "Keine", "Aus den Daten berechnen" oder "Wert".

- *Aus den Daten berechnen.* Schätzen Sie die natürliche Rücklaufquote aus den Stichprobendaten. Ihre Daten sollten einen Fall enthalten, der die Kontrollebene darstellt, für die der Wert der Kovariaten 0 ist. Probit schätzt die natürliche Ansprechrate anhand des Anteils der Antworten für die Kontrollstufe als Ausgangswert.
- *Wert.* Legt die natürliche Rücklaufquote im Modell fest (wählen Sie dieses Element aus, wenn Sie die natürliche Rücklaufquote im Voraus kennen). Geben Sie das natürliche Antwortverhältnis ein (das Verhältnis muss kleiner als 1 sein). Wenn die Antwort beispielsweise in 10 % der Zeit vorkommt, in der der Stimulus 0 ist, geben Sie den Wert 0,10 ein.

**Kriterien.** Hiermit können Sie Parameter für den Algorithmus zur iterativen Parameterschätzung festlegen. Sie können die Standardeinstellungen für die Maximalzahl der Iterationen, die Schrittweite und die Optimalitätstoleranz überschreiben.

## **Zusätzliche Funktionen beim Befehl PROBIT**

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Sie können eine Analyse beider Modelle anfordern, also des Probit-Modells und des Logit-Modells.
- Sie können festlegen, wie mit fehlenden Werten verfahren wird.
- Die Kovariaten können mit einer anderen Basis als Basis 10 oder dem natürlichen Logarithmus transformiert werden.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## **Quantilregression**

Regression ist eine statistische Methode, die im Wesentlichen bei der quantitativen Modellierung verwendet wird. Die mehrfache lineare Regression ist ein Basis- und Standardansatz, bei dem Forscher die Werte mehrerer Variablen verwenden, um die Mittelwerte eines Skalenergebnisses zu erklären oder



vorherzusagen. In vielen Fällen sind wir jedoch eher am Median oder an einem beliebigen Quantil des Skalenergebnisses interessiert.

Quantilregression modelliert die Beziehung zwischen einer Gruppe von Prädiktorvariablen (unabhängigen Variablen) und bestimmten Perzentilen (oder "Quantilen") einer Zielvariablen (abhängigen Variablen), die am häufigsten den Median aufweisen. Sie hat gegenüber der gewöhnlichen Regression nach der Methode der kleinsten Quadrate zwei wesentliche Vorteile:

- Die Quantilregression erstellt keine Annahmen über die Verteilung der Zielvariablen.
- Die Quantilregression ist gegenüber Ausreißern weniger empfindlich.

Die Quantilregression findet in der Forschung in Branchen wie Ökologie, Gesundheitswesen und Finanzökonomie breite Anwendung.

### **Beispiel:**

Was ist die Beziehung zwischen dem Haushaltsgesamteinkommen und dem Anteil des Einkommens, das für Lebensmittel ausgegeben wird? Das Gesetz von Engel ist eine Beobachtung in der Ökonomie, die besagt, dass bei steigendem Einkommen der Anteil des Einkommens, der für Nahrungsmittel ausgegeben wird, fällt, selbst wenn die absoluten Ausgaben für Nahrungsmittel steigen. Durch die Anwendung der Quantilregression auf diese Daten können Sie feststellen, welche Ausgaben für Lebensmittel unabhängig vom Mittelwert der Ausgaben für Lebensmittel 90 % der Familien abdecken (bei 100 Familien mit einem bestimmten Einkommen).

### **Statistik**

Quantilregression, Simplex-Ansatz, Frisch-Newton-Algorithmus für nicht lineare Optimierung innerer Punkte, Barrodale und Roberts, Bofinger, Hall, Sheather, Bandbreite, Signifikanzniveau, Matrizenmanipulation, Konvergenzkriterium, Regressionsgewichtungen, konstanter Term, vorhergesagtes Ziel, Vorhersageresiduen, Tabulierung, Vorhersagediagramme, Parameterschätzungen, Kovarianzmatrix, Korrelationsmatrix, beobachtete Werte, Konfidenzintervall

Diese Prozedur verwendet die Algorithmen, die von R. W. Koenker und G. W. Bassett vorgeschlagen wurden: Regression quantiles, *Econometrica*, 46, 33–50.

## **Erläuterungen der Daten für Quantilregression**

### **Daten**

Es ist eine einzige numerische abhängige Variable erforderlich. Die Zielvariable muss eine stetige Variable sein. Die Prädiktoren können stetige Variablen oder Dummy-Variablen für kategoriale Prädiktoren sein. Für die Ausführung einer Analyse ist entweder der konstante Term oder mindestens ein Prädiktor erforderlich.

### **Annahmen**

Quantilregression basiert nicht auf Annahmen zur Verteilung der Zielvariablen und unterliegt idR. nicht dem Einfluss von Randbeobachtungen.

### **Verwandte Prozeduren**

Die Quantilanalyse ist mit der gewöhnlichen Regression nach der Methode der einfachen kleinsten Quadrate verwandt.

## **Berechnen einer Quantilregressionsanalyse**

1. Wählen Sie in den Menüs Folgendes aus:

**Analysieren > Regression > Quantile...**

Im Dialogfeld können Sie die Ziel-, Faktor-, Kovariaten- und GewichtsvARIABLEN angeben, die für die Quantilregressionsanalyse verwendet werden sollen. Das Dialogfeld stellt außerdem Optionen zum Sparen von Speicher für komplexe Analysen oder große Datensets bereit.

2. Wählen Sie eine numerische Zielvariable aus. Zum Ausführen einer Analyse ist nur eine Zielvariable erforderlich. Es sind nur numerische Variablen zulässig.
3. Sie können bei Bedarf auch eine oder mehrere Faktorvariablen auswählen. Metrische Variablen sind nicht zulässig.

4. Sie können bei Bedarf auch eine oder mehrere Kovariatenvariablen auswählen. Zeichenfolgevariablen sind nicht zulässig.

**Anmerkung:** Wenn die Listen der **Faktoren** und **Kovariaten** leer sind und **Konstanter Term in Modell einschließen** im Dialogfeld "Modell" ausgewählt ist, wird die folgende Nachricht angezeigt:

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

5. Wählen Sie optional eine Regressionsgewichtungsvariable aus. Zeichenfolgevariablen sind nicht zulässig.
6. Wählen Sie optional **Hauptspeicher für komplexe Analyse oder große Datasets sparen** aus. Diese Einstellung steuert, ob die Daten während der Verarbeitung in einer externen Datei gespeichert werden oder nicht. Durch das Aktivieren dieser Einstellung lassen sich beim Ausführen von komplexen Analysen oder von Analysen mit großen Datasets Speicherressourcen freihalten.

## Quantilregression: Kriterien

Das Dialogfeld "Kriterien" enthält Optionen für

### Quantil

Stellt Optionen für die Angabe des Quantils bzw. der Quantilen bereit.

#### Einzelne Quantile angeben

Wird diese Option ausgewählt, ist mindestens ein Wert erforderlich, um die Analyse auszuführen. Es sind mehrere Werte zulässig und jeder Wert muss zu  $[0, 1]$  gehören. Sie können mehrere Werte mit jedem Wert angeben, die durch Leerzeichen getrennt sein müssen. Verwenden Sie die Schaltflächen **Hinzufügen**, **Ändern** und **Entfernen**, um mit den Werten in der Liste der Quantilwerte zu arbeiten.

Alle Werte müssen eindeutig sein (doppelte Werte sind nicht zulässig). Der Standardwert ist 0,5.

#### Rasterquantile angeben

Wird diese Option ausgewählt, kann ein Raster von Quantilen von einem **Startwert** (Wert1) bis zu einem **Endwert** (Wert2) mit dem Inkrement **By** (Wert3) angegeben werden. Wenn angegeben, ist nur eine einzige gültige Gruppe von  $[\text{Wert1 } T0 \text{ Wert2 } BY \text{ Wert3}]$  zulässig. Dabei muss  $0 \leq \text{Wert1} \leq \text{Wert2} \leq 1$  sein. Ist  $\text{Wert1} = \text{Wert2}$ , entspricht dies unabhängig von Wert3 der Angabe eines einzelnen Werts1.

### Schätzmethode

Stellt Optionen zum Angeben der Modellschätzmethode bereit.

#### Automatisch vom Programm ausgewählt

Ermöglicht die automatische Auswahl der geeigneten Schätzmethode. Hierbei handelt es sich um die Standardeinstellung.

#### Simplexalgorithmus

Ruft den von Barrodale und Roberts entwickelten Simplex-Algorithmus auf.

#### Frisch-Newton für nicht lineare Optimierung innerer Punkte

Ruft den Frisch-Newton-Algorithmus für nicht lineare Optimierung innerer Punkte auf.

### Nach der Schätzung

Stellt Optionen für die Nachschätzung der Varianz-Kovarianz der Parameterschätzungen und die Konfidenzintervall für die vorhergesagten Zielwerte bereit.

#### Annahme, dass Fälle IID sind

Wenn diese Einstellung ausgewählt ist, wird davon ausgegangen, dass die Fehlerterme unabhängig und identisch verteilt sind. Wenn die Einstellung nicht ausgewählt ist, kann die Berechnungszeit für große Modelle erheblich ansteigen. Diese Einstellung ist standardmäßig ausgewählt.

#### Bandbreitentyp

Bestimmt, welche Bandbreitenmethode verwendet wird, um die Varianz-Kovarianz-Matrix der Parameterschätzungen (**Bofinger** oder **Hall-Sheather**) zu schätzen. **Bofinger** ist die Standardeinstellung.

## Numerische Methode

Stellt die folgenden Optionen bereit:

### Singularitätstoleranz

Gibt den Toleranzwert für die Matrizenmanipulationen in der Methode innerer Punkte an. Der angegebene Wert muss ein einzelner Doppelwert sein ( $0, 10^{-3}$ ), wobei  $10^{-12}$  die Standardeinstellung ist.

### Konvergenz

Gibt das Konvergenzkriterium für die numerische Methode an. Der angegebene Wert muss ein einzelner Doppelwert sein ( $0, 10^{-3}$ ), wobei  $10^{-6}$  die Standardeinstellung ist.

### Maximalzahl der Iterationen

Gibt die maximale Anzahl der Iterationen an. Der angegebene Wert muss eine einzelne positive ganze Zahl sein. Der Standardwert ist 2000.

## Fehlende Werte

Stellt Optionen bereit, die festlegen, wie fehlende Werte behandelt werden.

### Benutzer- und systemdefiniert fehlende Werte ausschließen

Wenn diese Option ausgewählt ist, werden benutzer- und systemdefiniert fehlende Werte ausgeschlossen.

### Benutzerdefiniert fehlende Werte sind gültig

Wenn diese Option ausgewählt ist, werden benutzerdefiniert fehlende Werte als gültig behandelt.

## Konfidenzintervall (%)

Gibt das Signifikanzniveau an. Wenn diese Option angegeben ist, muss der Wert ein einzelner Doppelwert zwischen 0 und 100 sein. Der Standardwert ist 95.

## Quantilregression: Modell

Das Dialogfeld "Modell" enthält Optionen für die Angabe der Effekte und der Gewichtungen, die im Modell verwendet werden. Wurde es weggelassen oder automatisch festgelegt, dann enthält das Modell den konstanten Term und alle Haupteffekte mit den Kovariaten in der Liste der Kovariaten und den Faktoren in der Liste der Faktoren.

### Modelleffekte angeben

Das Standardmodell ist "Nur konstanter Term". Daher müssen Sie die anderen Modelleffekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme erstellen. Wenn **Terme erstellen** ausgewählt ist, sind für nicht verschachtelte Terme die folgenden Effekt- und Interaktionsoptionen verfügbar.

#### Haupteffekte

Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

#### Interaktion

Hiermit wird der Interaktionsterm mit der höchsten Ordnung für alle ausgewählten Variablen erzeugt.

#### Faktoriell

Erstellt alle möglichen Interaktionen und Haupteffekte der ausgewählten Variablen.

#### Alle 2-Wege

Hiermit werden alle möglichen Zweibegeinteraktionen der ausgewählten Variablen erstellt.

#### Alle 3-Wege

Hiermit werden alle möglichen Dreibegeinteraktionen der ausgewählten Variablen erzeugt.

#### Alle 4-Wege

Erstellt alle möglichen Vierbegeinteraktionen der ausgewählten Variablen.

#### Alle 5-Wege

Hiermit werden alle möglichen Fünfbegeinteraktionen der ausgewählten Variablen erzeugt.

Wenn **Verschachtelte Terme erstellen** ausgewählt ist, können Sie verschachtelte Terme erstellen. Verschachtelte Terme sind nützlich, um den Effekt eines Faktors oder einer Kovariate zu modellieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. So kann beispielsweise eine

Lebensmittelfilialkette den Ausgabengewohnheiten ihrer Kunden an mehreren Filialstandorten folgen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

Darüber hinaus können Sie Interaktionseffekte, wie polynomiale Terme mit derselben Kovariaten, einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

#### **Hinweise zum Verschachteln:**

- Um einen Effekt für eine Interaktion zwischen zwei Faktoren einzuschließen, verwenden Sie das Schlüsselwort BY oder den Stern (\*), um die Faktoren, die an der Interaktion beteiligt sind, zu verknüpfen.
- Faktoren innerhalb eines Interaktionseffekts müssen eindeutig sein.
- Verwenden Sie runde Klammern, um einen Effekt für das Verschachteln eines Terms innerhalb eines anderen Terms einzuschließen.
- Wenn mehrere Paare aus runden Klammern vorhanden sind, muss jedes Klammernpaar in einem anderen Klammernpaar eingeschlossen oder verschachtelt sein.
- Mehrfachverschachtelung ist zulässig.
- Interaktionen zwischen verschachtelten Effekten werden nicht unterstützt.

**Einschränkungen:** Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Interaktion müssen eindeutig sein. Dementsprechend ist die Angabe von  $A*A$  unzulässig, wenn  $A$  ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von  $A(A)$  unzulässig, wenn  $A$  ein Faktor ist.
- In einer Kovariate darf kein Effekt verschachtelt sein. Dementsprechend ist die Angabe von  $A(X)$  unzulässig, wenn  $A$  ein Faktor und  $X$  eine Kovariate ist.

#### **Konstanten Term in Modell einschließen**

Wenn diese Option ausgewählt ist, ist der konstante Term im Modell enthalten. Wenn diese Option nicht ausgewählt ist, ist zum Ausführen der Analyse mindestens ein Prädiktor erforderlich. Diese Einstellung ist standardmäßig aktiviert.

## **Quantilregression: Anzeigen**

Das Dialogfeld "Anzeigen" stellt Steuereinstellungen für Ausgabe und Darstellung bereit.

#### **Drucken**

Die folgenden Ausgabeoptionen sind verfügbar.

##### **Parameterschätzungen**

Zeigt Parameterschätzungen und die entsprechenden Teststatistiken und Konfidenzintervalle an. Wahlweise können Sie zusätzlich zu den rohen, unbearbeiteten Parameterschätzungen auch potenzierte Parameterschätzungen anzeigen.

##### **Kovarianzmatrix für Parameterschätzungen**

Zeigt die Kovarianzmatrix für die geschätzten Parameter an.

##### **Korrelationsmatrix für Parameterschätzungen**

Zeigt die Korrelationsmatrix für die geschätzten Parameter an.

#### **Darstellen und tabulieren**

Die folgenden Darstellungsoptionen sind verfügbar:

##### **Parameterschätzungen darstellen für**

Sie können auswählen, dass Parameterschätzungen für eine bestimmte Anzahl von oberen Effekten oder für alle Effekte in dem Modell dargestellt werden sollen. Über die Einstellung für die **anfänglichen xx Effekte** wird die Anzahl der Kategorien bzw. bei einem gemischten Effekt der Kategoriekombinationen gesteuert, die während der Interaktion zwischen einer Kovariaten oder ein oder zwei Faktoren dargestellt werden. Der Wert muss eine einzelne positive ganze Zahl sein (50 ist die Standardeinstellung).

### Hinweise:

- Vorhersagediagramme werden für alle Effekte erstellt, wenn der von Ihnen angegebene ganzzahlige Wert größer als die Anzahl der Kategorien oder Kombinationen ist.
- Die Einstellung ist nur gültig, wenn für die Einstellung **Quantilwerte** im Dialogfeld "Kriterien" mehrere Werte angegeben werden. Es werden keine Diagramme erstellt, wenn ein einzelner Quantilwert angegeben wird.

### Diagramm für vorhergesagt/beobachtet anzeigen

Steuert die Erstellung der vorhergesagten Werte im Vergleich zu den beobachteten Werten. Wenn diese Option aktiviert ist, wird ein einzelnes Diagramm erstellt, das die Punkte enthält (mit unterschiedlichen Farbpunkten, die unterschiedliche Quantile darstellen). Die Einstellung ist standardmäßig inaktiviert.

### Effekte im Modell vorhersagen

Wenn diese Option aktiviert ist, sind die folgenden Optionen verfügbar:

#### Oberste x Effekte darstellen oder tabulieren

Geben Sie die Anzahl der Top-Effekte an, deren Vorhersage-Diagramm oder Vorhersage-Tabelle erstellt werden soll. 3 ist der Standardwert.

**Anmerkung:** Die Vorhersagediagramme oder Vorhersagetabellen werden für alle Effekte erstellt, wenn der angegebene Wert größer als die Anzahl der gültigen Effekte im Modell ist.

#### Benutzerdefinierte Effekte darstellen oder tabulieren

Gültige Effektrichtlinien sind:

- Effekt mit einer Kovariate (einschließlich einer hohen Potenz der Kovariate selbst): Erstellen Sie ein einzelnes Diagramm, das die von verschiedenen Quantilen vorhergesagten Linien enthält.
- Effekt mit einem Faktor: Tabulieren Sie die Vorhersagen für die Kategorien des Faktors nach verschiedenen Quantilen.
- Effekt mit der Interaktion zweier Faktoren: Tabulieren Sie für jedes Quantil die Vorhersagen für die Kategorien zweier Faktoren.
- Effekt mit der Interaktion einer Kovariate und eines oder zweier Faktoren: Erstellen Sie für jedes Quantil ein Diagramm, das die Linien für jede Kategorie oder Kombination der Kategorien innerhalb des Effekts enthält.
- Die maximale Anzahl der darzustellenden Kombinationen wird über den Wert gesteuert, der als maximale Anzahl Kategorien bzw. in einem gemischten Effekt der Kategoriekombination festgelegt wird, die dargestellt werden können.

Effekte, die von **Modelleffekte** zur Liste **Vorhersagelinien** verschoben werden, werden für das Diagramm verwendet. In Fällen, in denen die angegebenen Effekte konstant sind (aus der Modellbildung entfernt), werden keine Diagramme erstellt.

Effekte, die aus der Liste **Modelleffekte** in die Liste **Vorhersagetabellen** verschoben wurden, werden für die Darstellung verwendet. In Fällen, in denen die angegebenen Effekte konstant sind (aus der Modellbildung entfernt), werden keine Tabellen erstellt.

#### Maximal xx Kategorien oder Kombinationen von Kategorien in einem gemischten Effekt darstellen

Steuert die maximale Anzahl der Kategoriekombinationen, die dargestellt werden sollen. Der Standardwert ist 10.

## Quantilregression: Speichern

Das Dialogfeld "Speichern" enthält Optionen für das Scoring (die Bewertung) des Modells.

### Vorhergesagter Wert der Antwort

Wenn diese Option ausgewählt ist, wird für den vorhergesagten Zielwert ein Scoring durchgeführt.

## Residuen

Bei Auswahl dieser Option werden die Vorhersageresiduen bewertet.

## Untergrenze des Vorhersageintervalls

Bei Auswahl dieser Option werden die Untergrenzen der Vorhersageintervalle bewertet.

## Obergrenze des Vorhersageintervalls

Wenn diese Option ausgewählt ist, werden die Obergrenzen der Vorhersageintervalle gescort.

**Anmerkung:** Für jede Speicheroption kann ein Variablenname angegeben werden. Wenn ein Stammmname angegeben wird, muss es sich um einen gültigen Variablennamen handeln. Ein Stammmname, gefolgt von einem Unterstrich "\_" und einem aussagekräftigen Quantilsuffix, wird verwendet, wenn mehrere Werte für die Einstellung **Quantilwerte** im Dialogfeld "Kriterien" angegeben werden.

## Quantilregression: Exportieren

Das Dialogfeld "Exportieren" stellt Optionen bereit, mit denen angegeben wird, welche Statistiken exportiert werden und wie dies erfolgen soll (externe Datendatei oder Datasets), und mit denen gesteuert wird, wie Daten während der Verarbeitung gehandhabt werden (normale Verarbeitung oder Speicherung in einer externen Arbeitsdatei während der Verarbeitung).

### Kovarianzmatrix der Parameterschätzungen

Wenn diese Option ausgewählt ist, werden Optionen zum Schreiben der Kovarianzmatrix der Parameterschätzungen in eine externe Datendatei oder ein zuvor deklariertes Dataset aktiviert.

### Korrelationsmatrix der Parameterschätzungen

Wenn diese Option ausgewählt ist, werden Optionen zum Schreiben der Korrelationsmatrix der Parameterschätzungen in eine externe Datendatei oder ein zuvor deklariertes Dataset aktiviert.

### Die Kovarianz-/Korrelationsmatrix wird in einem einzelnen Dataset oder einer externen Datei bei Vorhandensein mehrerer Regressionsquantile gespeichert

Wenn mehrere Quantile vorhanden sind, wird über diese Option festgelegt, ob Kovarianz-/Korrelationsmatrizen in einzelnen oder in mehreren Datasets oder externen Datendateien gespeichert werden. Wenn diese Option nicht aktiviert ist, werden Matrizen in einer einzelnen externen Datendatei oder einem Dataset gespeichert. Wenn diese Option aktiviert ist, werden Matrizen in mehreren externen Datendateien oder Datasets gespeichert. Die Einstellung wird nur wirksam, wenn mehrere Werte für die Einstellung **Quantilwerte** im Dialogfeld "Kriterien" angegeben werden.

**Anmerkung:** Diese Option ist nur verfügbar, wenn **Kovarianzmatrix der Parameterschätzungen** oder **Korrelationsmatrix der Parameterschätzungen** ausgewählt ist.

### Modellinformationen in XML-Datei exportieren

Wenn diese Option ausgewählt ist, können Sie Modellinformationen in eine XML-Datei in einer Speicherposition exportieren.

### Als XML exportieren

Wenn Sie **Modellinformation in XML-Datei exportieren** ausgewählt haben, können Sie auswählen, ob Parameterschätzungen und Kovarianzmatrizen oder nur Parameterschätzungen exportiert werden sollen. **Parameterschätzungen und Kovarianzmatrix** ist die Standardeinstellung.

## Dateinamenskonventionen

- Wenn ein einzelner Wert für die Einstellung **Quantilwerte** im Dialogfeld "Kriterien" angegeben wird, werden `savefile` und `dataset` verwendet, um die externe Datendatei oder das Dataset zu benennen.
- Wenn im Dialogfeld "Kriterien" mehrere Werte für die Einstellung **Quantilwerte** angegeben werden, wird jedes Quantil in einer externen Datendatei oder einem Dataset gespeichert.
- Ein Unterstrich "\_", gefolgt von einem aussagekräftigen Quantilsuffix, wird automatisch an den Namen der Datendatei oder des Dataset angehängt. Wenn beispielsweise 0,25, 0,50 und 0,75 als **Quantilwerte** angegeben wurden, wird das Suffix `_25`, `_50` bzw. `_75` an den Namen des entsprechenden Datasets (vor der Erweiterung `.sav`) angehängt.
- Zusätzliche Ziffern können für jedes Quantilsuffix angegeben werden (falls erforderlich).
- Die führende Null und das Dezimalzeichen in **Quantilwerte** werden im Suffix nicht verwendet.

- Wenn die wissenschaftliche Notation für **Quantilwerte** angegeben wurde, wird sie bei der Anzeige im Suffix in einen Dezimalwert konvertiert.

## Nicht lineare Regression

---

Die Nicht lineare Regression ist eine Methode, mit der Sie ein nicht lineares Modell für den Zusammenhang zwischen der abhängigen Variablen und einem Set von unabhängigen Variablen finden können. Im Gegensatz zur traditionellen linearen Regression, die auf die Schätzung linearer Modelle beschränkt ist, können Sie mit der nicht linearen Regression Modelle mit willkürlichen Beziehungen zwischen den abhängigen und unabhängigen Variablen schätzen. Dies wird durch den Einsatz eines iterativen Schätzalgorithmus erreicht. Beachten Sie, dass diese Prozedur für einfache Polynomialmodelle der Form  $Y = A + BX^{**2}$  nicht erforderlich ist. Durch die Definition von  $W = X^{**2}$  erhalten wir ein einfaches lineares Modell,  $Y = A + BW$ , das mit traditionellen Methoden wie dem linearen Regressionsverfahren geschätzt werden kann.

**Beispiel.** Kann die Populationsgröße in Abhängigkeit von der Zeit vorhergesagt werden? Anhand eines Streudiagramms können Sie sehen, dass ein enger Zusammenhang zwischen der Populationsgröße und der Zeit besteht. Da dieser Zusammenhang aber nicht linear ist, benötigen Sie hier die speziellen Schätzmethoden der Prozedur "Nicht lineare Regression". Durch eine geeignete Gleichung, wie beispielsweise ein logistisches Populationswachstumsmodell, kann eine gute Schätzung für das Modell erzielt werden. Dies wiederum ermöglicht Vorhersagen der Populationsgröße für Zeiträume, in denen diese nicht gemessen wurde.

**Statistik** Für jede Iteration: Parameterschätzungen und Residuenquadratsumme. Für jedes Modell: Quadratsumme für die Regression, Residuen, nicht korrigierter und korrigierter Gesamtwert, Parameterschätzungen, asymptotische Standardfehler und asymptotische Korrelationsmatrix der Parameterschätzungen.

**Anmerkung:** Eine nicht lineare Regression mit Nebenbedingungen verwendet die Algorithmen, die in NPSOL<sup>®</sup> (Gill, Murray, Saunders und Wright) zur Schätzung der Modellparameter vorgeschlagen und implementiert werden.

### Erläuterungen der Daten für die nicht lineare Regression

**Daten.** Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Kategoriale Variablen, wie beispielsweise Religion, Studienrichtung oder Wohnsitz, müssen in binäre (Dummy-)Variablen oder andere Typen von Kontrastvariablen umcodiert werden.

**Annahmen.** Die Ergebnisse sind nur dann gültig, wenn Sie eine Funktion festgelegt haben, die den Zusammenhang zwischen den abhängigen Variablen und den unabhängigen Variablen genau beschreibt. Außerdem ist die Auswahl der Anfangswerte sehr wichtig. Auch wenn Sie die korrekte Modellfunktion angegeben haben, kann es bei Verwendung von unpassenden Anfangswerten vorkommen, dass Ihr Modell nicht konvergiert oder statt einer global optimalen lediglich eine lokal optimale Lösung liefert.

**Verwandte Prozeduren.** Viele zunächst nicht linear erscheinende Modelle können in lineare Modelle transformiert werden, die mit der Prozedur "Lineare Regression" analysiert werden können. Wenn Sie sich nicht sicher sind, welches das geeignete Modell ist, können Sie mit der Prozedur "Kurvenanpassung" nützliche funktionale Zusammenhänge in Ihren Daten ermitteln.

### Berechnen einer nicht linearen Regression

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Regression > Nicht linear...**

2. Wählen Sie eine numerische abhängige Variable aus der Liste der Variablen in Ihrem aktiven Dataset.

3. Sie erstellen eine Modellformel, indem Sie einen Ausdruck in das Feld **Modellformel** eingeben oder die Komponenten (Variablen, Parameter, Funktionen) in das Feld einfügen.

4. Klicken Sie auf **Parameter**, um die Parameter in Ihrem Modell festzulegen.

Ein segmentiertes Modell nimmt verschiedene Formen innerhalb verschiedener Abschnitte des Definitionsbereiches an. Sie müssen Bedingungslogik verwenden, um es in einer einzelnen Modellformel auszudrücken.

## Bedingungslogik (nicht lineare Regression)

Sie können ein segmentiertes Modell festlegen, indem Sie Bedingungslogik verwenden. Sie verwenden Bedingungslogik in einer Modellformel oder einer Verlustfunktion, indem Sie für jede Bedingung die Summe einer Reihe von Termen bilden. Jeder Term besteht aus dem Produkt eines in Klammern stehenden logischen Ausdrucks und dem Ausdruck, der sich ergibt, wenn der logische Ausdruck wahr ist.

Betrachten Sie zum Beispiel ein segmentiertes Modell, bei dem  $0$  für  $X \leq 0$ ,  $X$  für  $0 < X < 1$ , und  $1$  für  $X \geq 1$  entspricht. Der Ausdruck hierfür ist:

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

Die logischen Ausdrücke in Klammern nehmen entweder den Wert  $1$  (wahr) oder den Wert  $0$  (falsch) an. Also:

Bei  $X \leq 0$  ergibt sich  $1 * 0 + 0 * X + 0 * 1 = 0$ .

Bei  $0 < X < 1$  ergibt sich  $0 * 0 + 1 * X + 0 * 1 = X$ .

Wenn  $X \geq 1$  ist, wird auf  $0 * 0 + 0 * X + 1 * 1 = 1$  reduziert.

Kompliziertere Beispiele können leicht gebildet werden, indem andere logische Ausdrücke und Ergebnisausdrücke eingesetzt werden. Denken Sie an doppelte Ungleichheiten wie beispielsweise  $0 < X < 1$ , must be written as compound expressions, such as  $(X > 0 \ \& \ X < 1)$ .

In logischen Ausdrücken können auch Zeichenfolgevariablen verwendet werden:

$$(\text{Stadt} = \text{'Frankfurt'}) * \text{lebkost} + (\text{Stadt} = \text{'Regensburg'}) * 0,59 * \text{lebkost}$$

Damit erhalten Sie zwei Ausdrücke: den Ausdruck für die Einwohner Frankfurts, der dem Wert der Variablen *Lebkost* entspricht, und den Ausdruck für die Einwohner Regensburgs, dessen Wert 59 % des ersten Werts beträgt. Zeichenfolgekonstanten müssen, wie hier gezeigt, in Anführungszeichen oder Apostrophe eingeschlossen werden.

## Nicht lineare Regression: Parameter

Die Parameter sind Teil Ihres Modells, das die Prozedur "Nicht lineare Regression" schätzt. Zu den Parametern zählen additive Konstanten, multiplikative Koeffizienten, Exponenten oder Werte, die beim Auswerten von Funktionen verwendet werden. Alle von Ihnen definierten Parameter werden mit ihren Anfangswerten in der Parameterliste des Hauptdialogfelds angezeigt.

**Name.** Sie müssen für jeden Parameter einen Namen festlegen. Dieser Name muss ein gültiger Variablenname sein und mit dem Namen übereinstimmen, der in der Modellformel im Hauptdialogfeld verwendet wird.

**Anfangswert.** Hier können Sie den Anfangswert des Parameters festlegen. Dieser sollte so weit wie möglich der erwarteten endgültigen Lösung entsprechen. Unpassende Anfangswerte können dazu führen, dass im Ergebnis keine Konvergenz erreicht wird, die Lösung nur lokal und nicht global gültig ist oder die Lösung physisch unmöglich ist.

**Anfangswerte aus voriger Analyse verwenden.** Falls Sie bereits eine nicht lineare Regression in diesem Dialogfeld durchgeführt haben, können Sie diese Option wählen, um als Anfangswerte der Parameter die Werte einer früheren Analyse zu übernehmen. Damit haben sie die Möglichkeit, mit der Suche fortzufahren, wenn der Algorithmus langsam konvergiert. (Die ursprünglich gewählten Anfangswerte werden aber immer noch in der Parameterliste des Hauptdialogfelds angezeigt.)

*Hinweis:* Diese Option in diesem Dialogfeld bleibt bis zum Ende Ihrer Sitzung ausgewählt. Falls Sie das Modell ändern, vergessen Sie nicht, die Option abzuwählen.



## Nicht lineare Regression: Gängige Modelle

In der nachfolgenden Tabelle werden Syntaxbeispiele für viele publizierte nicht lineare Regressionsmodelle aufgeführt. Ein zufällig gewähltes Modell wird für Ihre Daten fast immer ungeeignet sein. Sie benötigen geeignete Anfangswerte für die Parameter und bei einigen Modellen sind Nebenbedingungen erforderlich, damit diese konvergieren.

Tabelle 1. Beispiele für Modellformeln:

| Name                                       | Modellformel   |
|--|--|
| Asymptotische Regression                   | $b_1 + b_2 \cdot \text{Exp}(b_3 \cdot x)$  |
| Asymptotische Regression                   | $b_1 - (b_2 \cdot (b_3^{**} x))$   |
| Dichte                                     | $(b_1 + b_2 \cdot x)^{**} (-1 / b_3)$  |
| Gauß                                       | $b_1 \cdot (1 - b_3 \cdot \text{Exp}(-b_2 \cdot x^{**} 2))$                                  |
| Gompertz                                   | $b_1 \cdot \text{Exp}(-b_2 \cdot \text{Exp}(-b_3 \cdot x))$                                  |
| Johnson-Schumacher                         | $b_1 \cdot \text{Exp}(-b_2 / (x + b_3))$   |
| Log-modifiziert                            | $(b_1 + b_3 \cdot x)^{**} b_2$   |
| Log-logistisch                             | $b_1 - \ln(1 + b_2 \cdot \text{Exp}(-b_3 \cdot x))$  |
| Metcherlich-Gesetz der abnehmenden Erträge | $b_1 + b_2 \cdot \text{Exp}(-b_3 \cdot x)$   |
| Michaelis-Menten                           | $b_1 \cdot x / (x + b_2)$  |
| Morgan-Mercer-Florin                       | $(b_1 \cdot b_2 + b_3 \cdot x^{**} b_4) / (b_2 + x^{**} b_4)$                                |
| Peal-Reed                                  | $b_1 / (1 + b_2 \cdot \text{Exp}(-(b_3 \cdot x + b_4 \cdot x^{**} 2 + b_5 \cdot x^{**} 3)))$ |
| Verhältnis der 3. Potenzen                 | $(b_1 + b_2 \cdot x + b_3 \cdot x^{**} 2 + b_4 \cdot x^{**} 3) / (b_5 \cdot x^{**} 3)$       |
| Verhältnis der Quadrate                    | $(b_1 + b_2 \cdot x + b_3 \cdot x^{**} 2) / (b_4 \cdot x^{**} 2)$                            |
| Richards                                   | $b_1 / ((1 + b_3 \cdot \text{Exp}(-b_2 \cdot x))^{**} (1 / b_4))$                            |
| Verhulst                                   | $b_1 / (1 + b_3 \cdot \text{Exp}(-b_2 \cdot x))$   |
| Von Bertalanffy                            | $(b_1^{**} (1 - b_4) - b_2 \cdot \text{Exp}(-b_3 \cdot x))^{**} (1 / (1 - b_4))$             |
| Weibull (PP Plots Test Dist.)              | $b_1 - b_2 \cdot \text{Exp}(-b_3 \cdot x^{**} b_4)$  |
| Ertragsdichte                              | $(b_1 + b_2 \cdot x + b_3 \cdot x^{**} 2)^{**} (-1)$   |

## Nicht lineare Regression: Verlustfunktion

Die **Verlustfunktion** in der nicht linearen Regression ist die vom Algorithmus minimierte Funktion. Wählen Sie entweder **Summe der quadrierten Residuen** aus, um die Summe der quadrierten Residuen zu minimieren oder wählen Sie **Benutzerdefinierte Verlustfunktion** aus, um eine andere Funktion zu minimieren.

Wenn Sie **Benutzerdefinierte Verlustfunktion** auswählen, müssen Sie die Verlustfunktion definieren. Die Summe (über alle Fälle) dieser Verlustfunktion sollte durch die Auswahl der Parameterwerte minimiert werden.

- Die meisten Verlustfunktionen enthalten die spezielle Variable *RESID\_* (das Residuum). (Sie können die Standardverlustfunktion "Summe der quadrierten Residuen" explizit als *RESID\_\*\*2* eingeben.) Wenn Sie den vorhergesagten Wert in Ihrer Verlustfunktion verwenden müssen, können Sie diesen einfach berechnen, indem Sie das Residuum von der abhängigen Variablen abziehen.
- Mit Bedingungslogik können Sie eine bedingte Verlustfunktion festlegen.

Sie können einen Ausdruck in das Feld "Benutzerdefinierte Verlustfunktion" eingeben oder die Komponenten des Ausdrucks in das Feld einfügen. Zeichenfolgekennstanten müssen in Anführungszeichen oder

Apostrophe eingeschlossen werden. Das Format der numerischen Konstanten muss amerikanisch sein. Sie müssen als Dezimaltrennzeichen also einen Punkt verwenden.

## Nicht lineare Regression: Nebenbedingungen für Parameter

Eine **Nebenbedingung** ist eine Einschränkung der zulässigen Werte für einen Parameter bei der iterativen Lösungssuche. Lineare Ausdrücke werden berechnet, bevor ein Schritt ausgeführt wird. Sie können also lineare Nebenbedingungen verwenden, um Schritte zu vermeiden, die zu einem Überlauf führen könnten. Nicht lineare Ausdrücke werden berechnet, nachdem ein Schritt ausgeführt wurde.

Für jede Gleichung oder Ungleichung sind die folgenden Elemente erforderlich:

- Einen Ausdruck, der mindestens einen Parameter des Modells enthält. Geben Sie den Ausdruck ein oder verwenden Sie den numerischen Tastenblock, mit dem Sie Zahlen, Operatoren oder Klammern in den Ausdruck einfügen können. Sie können den bzw. die benötigten Parameter entweder zusammen mit dem Rest des Ausdrucks eingeben oder aus der Parameterliste auf der linken Seite einfügen. Gewöhnliche Variablen können in den Nebenbedingungen nicht verwendet werden.
- Einer der drei logischen Operatoren  $\leq$ ,  $=$ , or  $\geq$ .
- Eine numerische Konstante, mit welcher der Ausdruck unter Verwendung des logischen Operators verglichen wird. Geben Sie die Konstante ein. Numerische Konstanten müssen im amerikanischen Format mit einem Punkt als Dezimaltrennzeichen eingegeben werden.

## Nicht lineare Regression: Neue Variablen speichern

Sie können eine Anzahl neuer Variablen in der aktiven Datendatei speichern. Die folgenden Optionen sind verfügbar: "Residuen", "Vorhergesagte Werte", "Ableitungen" und "Werte der Verlustfunktion". Diese Variablen können in nachfolgenden Analysen verwendet werden, um die Anpassungsgüte des Modells zu testen oder Problemfälle zu identifizieren.

- *Residuals*. Speichert Residuen unter dem Variablennamen "resid".
- *Vorhergesagte Werte*. Speichert vorhergesagte Werte unter dem Variablennamen "pred\_".
- *Ableitungen*. Für jeden Modellparameter wird eine Ableitung gespeichert. Ableitungsnamen werden durch das Präfix "d" erstellt. auf die ersten sechs Zeichen von Parameternamen.
- *Werte der Verlustfunktion*. Diese Option ist verfügbar, wenn Sie Ihre eigene Verlustfunktion angeben. Den Werten der Verlustfunktion wird der Variablenname "loss\_" zugewiesen.

## Nicht lineare Regression: Optionen

Mit den folgenden Optionen können Sie die Einstellungen verschiedener Aspekte der nicht linearen Regressionsanalyse vornehmen:

**Bootstrap-Schätzungen.** Eine Methode zur Schätzung des Standardfehlers einer Statistik mithilfe wiederholter Stichproben aus dem ursprünglichen Dataset. Dabei wird Ziehen mit Zurücklegen verwendet, um viele Stichproben zu erhalten, die denselben Umfang haben wie das Originaldataset. Für jede dieser Stichproben wird die nicht lineare Gleichung geschätzt. Der Standardfehler jeder Parameterschätzung wird dann als Standardabweichung der Bootstrap-Schätzungen berechnet. Parameterwerte aus den ursprünglichen Daten werden als Anfangswerte für die einzelne Bootstrap-Stichprobe verwendet. Dies macht den Algorithmus für sequenzielle quadratische Programmierung erforderlich.

**Schätzmethode.** Hiermit können Sie, soweit möglich, eine Schätzmethode auswählen. (Bei einigen Optionen dieser oder anderer Dialogfelder ist der Algorithmus für die sequenzielle quadratische Optimierung erforderlich.) Verfügbar sind "Sequenzielle quadratische Optimierung" und "Levenberg-Marquardt".

- *Sequenzielle quadratische Optimierung.* Diese Methode ist für Modelle mit und ohne Bedingungen verfügbar. Eine sequenzielle quadratische Optimierung wird automatisch verwendet, wenn Sie ein Modell mit Nebenbedingungen, eine benutzerdefinierte Verlustfunktion oder Bootstrapping festlegen. Sie können neue Werte für die Maximalzahl an Iterationen und die Schrittweite angeben und Sie können die Auswahl in den Dropdown-Listen für Optimalitätstoleranz, Funktionsgenauigkeit und unendliche Schrittweite ändern.

- *Levenberg-Marquardt*. Dies ist der Standardalgorithmus für Modelle ohne Bedingungen. Das Levenberg-Marquardt-Verfahren steht nicht zur Verfügung, wenn Sie ein Modell mit Nebenbedingungen, eine benutzerdefinierte Verlustfunktion oder Bootstrapping festlegen. Sie können neue Werte für die Maximalzahl an Iterationen angeben und Sie können die Auswahl in den Dropdown-Listen für Quadratsummenkonvergenz und Parameterkonvergenz ändern.

## Interpretieren von Ergebnissen der nicht linearen Regression

Probleme bei der nicht linearen Regression deuten oft auf Schwierigkeiten bei der Berechnung hin:

- Die Anfangswerte der Parameter beeinflussen die Konvergenz. Versuchen Sie, geeignete Anfangswerte auszuwählen, die so weit wie möglich der erwarteten endgültigen Lösung entsprechen.
- Manchmal funktioniert bei einem bestimmten Problem ein Algorithmus besser als ein anderer. Wählen Sie im Dialogfeld "Optionen" einen anderen Algorithmus, wenn dieser verfügbar ist. (Wenn Sie eine Verlustfunktion oder bestimmte Arten von Nebenbedingungen festlegen, können Sie den Levenberg-Marquardt-Algorithmus nicht verwenden.)
- Wenn die Iteration lediglich deshalb stoppt, weil die maximale Anzahl der Iterationen erreicht ist, so ist das "endgültige" Modell möglicherweise keine gute Lösung. Wählen Sie im Dialogfeld "Nicht lineare Regression: Parameter" die Option **Anfangswerte aus voriger Analyse verwenden** aus, um mit der Iteration fortzufahren. Noch besser ist es, andere Anfangswerte zu wählen.
- Modelle, die die Potenzierung von oder mit großen Datenwerten erfordern, können einen Über- oder Unterlauf verursachen (also Zahlen, die zu groß oder zu klein sind, als dass sie auf dem Computer dargestellt werden könnten). Sie können dies manchmal durch die geeignete Auswahl von Anfangswerten oder durch Aufstellen von Nebenbedingungen für die Parameter verhindern.

## Zusätzliche Funktionen beim Befehl NLR

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Sie können eine Datei angeben, aus der die Anfangswerte für die Parameterschätzungen gelesen werden.
- Es können mehr als eine Modellfunktion und mehr als eine Verlustfunktionen festgelegt werden. Dies erleichtert die Verwendung eines segmentierten Modells.
- Sie können Ihre eigenen Ableitungen verwenden, anstatt die vom Programm berechneten zu verwenden.
- Die Anzahl der zu berechnenden Bootstrap-Stichproben kann festgelegt werden.
- Es können zusätzliche Iterationskriterien festgelegt werden. Dies schließt das Festsetzen eines kritischen Werts für die Überprüfung der Ableitung und das Definieren eines Konvergenzkriteriums für die Korrelationen zwischen den Residuen und den Ableitungen ein.

Mit den zusätzlichen Kriterien für den Befehl CNLR (nicht lineare Regression mit Nebenbedingungen) verfügen Sie über die folgenden Möglichkeiten:

- Sie können die maximale Anzahl der sekundären Iterationen festlegen, die innerhalb jeder Hauptiteration zulässig sind.
- Für das Prüfen der Ableitung kann ein kritischer Wert festgelegt werden.
- Eine Schrittweite kann festgelegt werden.
- Sie können eine Crash-Toleranz festlegen, um zu bestimmen, ob sich die Anfangswerte innerhalb ihrer festgelegten Grenzen befinden.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Gewichtungsschätzung

Standardmodelle für die lineare Regression gehen davon aus, dass die Varianz in der untersuchten Grundgesamtheit konstant ist. Wenn dies nicht der Fall ist (zum Beispiel, wenn Fälle mit hohen Werten bei einem bestimmten Attribut eine größere Variabilität aufweisen als Fälle mit niedrigen Werten für

dieses Attribut), liefert die lineare Regression unter Verwendung der gewöhnlichen kleinsten Quadrate (OLS) keine optimalen Modellschätzungen. Wenn die Differenzen in der Variabilität von einer anderen Variablen vorhergesagt werden können, dann können die Koeffizienten eines linearen Regressionsmodells mit der Prozedur "Gewichtungsschätzung" unter Verwendung von gewichteten kleinsten Quadraten (WLS) berechnet werden. Dabei wird den genaueren Beobachtungen (jene mit kleinerer Variabilität) ein größeres Gewicht beim Bestimmen der Regressionskoeffizienten zugewiesen. Mit der Prozedur "Gewichtungsschätzung" werden eine Reihe von Gewichtungstransformationen getestet, und es wird angezeigt, welche die beste Anpassung an die Daten ergibt.

**Beispiel.** Welche Auswirkung haben Inflation und Arbeitslosigkeit auf die Börsenkurse? Da Aktien mit höheren Aktienwerten oft eine größere Variabilität als jene mit niedrigen Aktienwerten aufweisen, liefern gewöhnliche kleinste Quadrate keine optimalen Schätzungen. Mit der Prozedur "Gewichtungsschätzung" können die Auswirkungen des Aktienpreises auf die Variabilität von Preisänderungen beim Berechnen des linearen Modells berücksichtigt werden.

**Statistik** Log-Likelihood-Werte für jeden Exponenten der getesteten Quellenvariablen für die Gewichtung, multiples  $R$ ,  $R$ -Quadrat, korrigiertes  $R$ -Quadrat, ANOVA-Tabelle für das WLS-Modell, nicht standardisierte und standardisierte Parameterschätzungen und Log-Likelihood für das WLS-Modell.

## Erläuterungen der Daten für die Gewichtungsschätzung

**Daten.** Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Kategoriale Variablen, wie beispielsweise Religion, Studienrichtung oder Wohnsitz, müssen in binäre (Dummy-)Variablen oder andere Typen von Kontrastvariablen umcodiert werden. Die Gewichtungsvariable sollte quantitativ sein und mit der Variabilität in der abhängigen Variablen in Beziehung stehen.

**Annahmen.** Für jeden Wert der unabhängigen Variablen muss die abhängige Variable normalverteilt vorliegen. Die Beziehung zwischen der abhängigen Variablen und allen unabhängigen Variablen sollte linear sein, und alle Beobachtungen sollten unabhängig sein. Die Varianz der abhängigen Variablen kann für verschiedene Stufen der unabhängigen Variablen unterschiedlich sein, aber die Differenzen müssen auf der Grundlage der Gewichtungsvariablen vorhersagbar sein.

**Verwandte Prozeduren.** Sie können die Prozedur "Explorative Datenanalyse" zum Sichten der Daten verwenden. In der Prozedur "Explorative Datenanalyse" werden Tests auf Normalverteilung und Homogenität der Varianzen sowie grafische Darstellungen zur Verfügung gestellt. Falls in der abhängigen Variablen für alle Stufen der unabhängigen Variablen offenbar gleiche Varianzen vorliegen, können Sie die Prozedur "Lineare Regression" verwenden. Wenn die Daten eine der Annahmen verletzen (wie zum Beispiel die Normalverteilung), versuchen Sie, diese zu transformieren. Wenn zwischen Ihren Daten kein linearer Zusammenhang besteht und eine Transformation erfolglos bleibt, sollten Sie ein alternatives Modell in der Prozedur "Kurvenanpassung" verwenden. Wenn die abhängige Variable dichotom ist, zum Beispiel, ob ein bestimmter Verkauf abgeschlossen oder ein Teil defekt ist, verwenden Sie die Prozedur "Logistische Regression". Wenn die abhängige Variable zensiert ist, zum Beispiel die postoperative Überlebenszeit, verwenden Sie Sterbetafeln, Kaplan-Meier oder die Cox-Regression (verfügbar in "Benutzerdefinierte Tabellen" und "Erweiterte Statistikfunktionen"). Wenn Ihre Daten nicht unabhängig sind, weil Sie beispielsweise dieselbe Person unter verschiedenen Bedingungen beobachten, verwenden Sie die Prozedur "Messwiederholungen". Diese steht Ihnen in "Benutzerdefinierte Tabellen" und "Erweiterte Statistikfunktionen" zur Verfügung.

## Berechnen einer Analyse zur Gewichtungsschätzung

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Regression > Gewichtungsschätzung...**

2. Wählen Sie eine abhängige Variable aus.

3. Wählen Sie mindestens eine unabhängige Variable aus.

4. Wählen Sie die Variable als Gewichtungsvariable aus, welche die Quelle der Heteroskedastizität ist.

### **Gewichtungsvariable**

Die Daten werden durch den reziproken Wert dieser Variablen gewichtet, der hoch potenziert wird. Die Regressionsgleichung wird für jeden Wert des Exponenten in einem angegebenen Bereich berechnet und gibt den Exponenten an, der die Log-Likelihood-Funktion maximiert.

### **Leistungsbereich**

Wird in Verbindung mit der GewichtungsvARIABLEN verwendet, um Gewichtungen zu berechnen. Für jeden Wert im Exponentenbereich wird jeweils eine Regressionsgleichung angepasst. Die Werte, die in den Feldern "Exponentenbereich" und "bis" eingegeben werden, müssen im Bereich von -6,5 bis 7,5 liegen. Der Exponentenbereich reicht vom Minimum bis zum Maximum, in Inkrementen, die durch den angegebenen Wert bestimmt sind. Die Gesamtanzahl der Werte im Exponentenbereich ist auf 150 begrenzt.

## **Gewichtungsschätzung: Optionen**

Sie können Optionen für die Analyse zur Gewichtungsschätzung festlegen.

**Beste Gewichtung als neue Variable speichern.** Hiermit fügen Sie der aktiven Datei die GewichtungsvARIABLEN hinzu. Diese Variable heißt *WGT\_n*, wobei *n* eine Nummer zur eindeutigen Kennzeichnung der Variablen ist.

**ANOVA und Schätzungen anzeigen.** Hiermit legen Sie fest, wie die Statistiken in der Ausgabe angezeigt werden. Sie können zwischen "Für besten Exponenten" und "Für jeden Exponenten" auswählen.

## **Zusätzliche Funktionen beim Befehl WLS**

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Sie können einen einzelnen Wert für den Exponenten angeben.
- Sie können eine Liste mit Exponentenwerten festlegen oder einen Wertebereich mit einer Liste von Werten für den Exponenten kombinieren.

Vollständige Syntaxinformationen finden Sie unter *Befehlssyntaxreferenz*.

## **Zweistufige Regressionsmethode der kleinsten Quadrate**

---

Standardmodelle für die lineare Regression gehen davon aus, dass Fehler in der abhängigen Variablen nicht mit der/den unabhängigen Variablen korrelieren. Ist dies nicht der Fall (zum Beispiel, wenn zwischen den Variablen eine Interaktion besteht), dann liefert die lineare Regression unter Verwendung von gewöhnlichen kleinsten Quadraten (OLS - Ordinary Least Squares) keine optimalen Modellschätzungen mehr. Die zweistufige Regressionsmethode der kleinsten Quadrate verwendet Instrumentalvariablen, die nicht mit den Fehlertermen korrelieren. Damit werden in der ersten Stufe Näherungswerte für die problematischen Prädiktoren berechnet, mit denen in der zweiten Stufe ein lineares Regressionsmodell der abhängigen Variablen ermittelt wird. Da die berechneten Werte auf Variablen basieren, die nicht mit den Fehlern korrelieren, sind die Ergebnisse des zweistufigen Modells optimal.

**Beispiel.** Steht die Nachfrage nach einer Ware in einer Beziehung zu ihrem Preis und zum Einkommen der Konsumenten? Die Schwierigkeit in diesem Modell besteht darin, dass sich Preis und Nachfrage in reziproker Weise beeinflussen. Das heißt, der Preis kann die Nachfrage beeinflussen und umgekehrt. Die zweistufige Regressionsmethode der kleinsten Quadrate könnte das Einkommen der Konsumenten und den zeitversetzt entstehenden Preis nutzen, um einen Näherungswert für den Preis zu errechnen, der nicht mit den Messfehlern für die Nachfrage korreliert. Dieser Näherungswert wird in dem ursprünglich angegebenen Modell anstelle des Preises eingesetzt. Anschließend wird es neu geschätzt.

**Statistiken.** Für jedes Modell: standardisierte und nicht standardisierte Regressionskoeffizienten, multiples  $R$ ,  $R^2$ , korrigiertes  $R^2$ , Standardfehler der Schätzung, Tabelle für die Varianzanalyse, vorhergesagte Werte und Residuen. Weiterhin werden 95-%-Konfidenzintervalle für jeden Regressionskoeffizienten sowie Korrelations- und Kovarianzmatrizen der Parameterschätzungen angezeigt.

## Erläuterungen der Daten für die zweistufige Regressionsmethode der kleinsten Quadrate

**Daten.** Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Kategoriale Variablen, wie beispielsweise Religion, Studienrichtung oder Wohnsitz, müssen in binäre (Dummy-)Variablen oder andere Typen von Kontrastvariablen umcodiert werden. *Endogene* erklärende Variablen müssen quantitativ sein (nicht kategorial).

**Annahmen.** Für jeden Wert der unabhängigen Variablen muss die abhängige Variable normalverteilt vorliegen. Die Varianz der Verteilung der abhängigen Variablen muss für alle Werte der unabhängigen Variablen konstant sein. Zwischen der abhängigen und jeder unabhängigen Variablen muss ein linearer Zusammenhang bestehen.

**Verwandte Prozeduren.** Wenn Sie annehmen, dass keine der Vorhersagevariablen mit den Fehlern in der abhängigen Variablen korreliert, dann können Sie die Prozedur "Lineare Regression" verwenden. Wenn die Daten eine der Voraussetzungen zu verletzen scheinen (beispielsweise Normalverteilung oder konstante Varianz), dann versuchen Sie, die Daten zu transformieren. Wenn zwischen Ihren Daten kein linearer Zusammenhang besteht und eine Transformation erfolglos bleibt, sollten Sie ein alternatives Modell in der Prozedur "Kurvenanpassung" verwenden. Wenn sich die abhängige Variable dichotom verhält, beispielsweise ein bestimmter Verkauf abgeschlossen ist oder nicht, dann verwenden Sie die Prozedur "Logistische Regression". Wenn Ihre Daten nicht unabhängig sind, weil Sie beispielsweise dieselbe Person unter verschiedenen Bedingungen beobachten, verwenden Sie die Prozedur "Messwiederholungen".

## Berechnen einer Analyse nach der zweistufigen Regressionsmethode der kleinsten Quadrate

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Regression > Zweistufige Regressionsmethode der kleinsten Quadrate...**

2. Wählen Sie eine abhängige Variable aus.

3. Wählen Sie mindestens eine erklärende Variable (Prädiktor) aus.

4. Wählen Sie mindestens eine Instrumentalvariable aus.

- *Instrumentalvariablen.* Diese Variablen werden verwendet, um die vorhergesagten Werte für die endogenen Variablen in der ersten Stufe der zweistufigen Analyse der kleinsten Quadrate zu berechnen. Dieselben Variablen können sowohl in der Liste der erklärenden Variablen als auch in der Liste der Instrumentalvariablen erscheinen. Die Anzahl der Instrumentalvariablen muss mindestens der Anzahl der erklärenden Variablen entsprechen. Wenn alle erklärenden und Instrumentalvariablen in den Listen gleich sind, entsprechen die Ergebnisse den Ergebnissen der linearen Regression.

Erklärende Variablen, die nicht als Instrumentalvariablen festgelegt sind, werden als endogen betrachtet. Normalerweise werden alle exogenen Variablen in der Liste "Erklärende Var." ebenfalls als Instrumentalvariablen festgelegt.

## Zweistufige kleinste Quadrate: Optionen

Sie können eine der folgenden Optionen für Ihre Analyse auswählen:

**Neue Variablen speichern.** Hiermit können Sie Ihrer aktiven Datei neue Variablen hinzufügen. Die verfügbaren Optionen sind "Vorhersagewert" und "Residuen".

**Kovarianz der Parameter anzeigen.** Hiermit können Sie sich die Kovarianzmatrix für die Parameterschätzungen anzeigen lassen.

## Zusätzliche Funktionen beim Befehl 2SLS

Mit der Befehlssyntaxsprache können Sie mehrere Gleichungen gleichzeitig schätzen. Vollständige Informationen zur Syntax finden Sie in der *Befehlssyntaxreferenz*.

## Codierungsschemata für kategoriale Variablen

In vielen Prozeduren können Sie festlegen, dass eine kategoriale unabhängige Variable automatisch durch ein Set Kontrastvariablen ersetzt wird, die dann gemeinsam als Block in eine Gleichung aufgenommen oder aus ihr entfernt werden. Sie können festlegen, wie das Set der Kontrastvariablen codiert werden soll (in der Regel im Unterbefehl CONTRAST). Dieser Anhang erläutert die verschiedenen Kontrasttypen des Unterbefehls CONTRAST.

### Abweichung

**Abweichung vom Gesamtmittelwert.** Ausgedrückt als Matrix weisen diese Kontraste die folgende Form auf:

```
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)
```

Hierbei entspricht  $k$  der Anzahl der Kategorien in der unabhängigen Variablen, und die letzte Kategorie wird in der Standardeinstellung weggelassen. Die Abweichungskontraste für eine unabhängige Variable mit drei Kategorien lauten beispielsweise wie folgt:

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

Wenn Sie nicht die letzte, sondern eine andere Kategorie weglassen möchten, geben Sie die Nummer der wegzulassenden Kategorie nach dem Schlüsselwort DEVIATION in Klammern an. Mit dem folgenden Unterbefehl werden beispielsweise die Abweichungen für die erste und dritte Kategorie berechnet, und die zweite Kategorie wird weggelassen:

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

Wenn *Faktor* drei Kategorien aufweist, wird die folgende Kontrastmatrix berechnet:

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

### Einfach

**Einfache Kontraste.** Hierbei wird jede Faktorstufe mit der vorausgehenden verglichen. Die allgemeine Matrixform lautet:

```
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)
```

Hierbei entspricht  $k$  der Anzahl der Kategorien in der unabhängigen Variablen. Die einfachen Kontraste für eine unabhängige Variable mit vier Kategorien lauten beispielsweise wie folgt:

```
(1/4  1/4  1/4  1/4)
( 1  0  0 -1)
( 0  1  0 -1)
( 0  0  1 -1)
```

Wenn Sie nicht die letzte, sondern eine andere Kategorie als Referenzkategorie verwenden möchten, geben Sie die laufende Nummer der Referenzkategorie nach dem Schlüsselwort SIMPLE in Klammern ein. Diese Nummer entspricht nicht notwendigerweise dem Wert dieser Kategorie. Mit dem folgenden Unterbefehl CONTRAST wird beispielsweise eine Kontrastmatrix berechnet, bei der die zweite Kategorie weggelassen wird:

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

Wenn *Faktor* vier Kategorien aufweist, wird die folgende Kontrastmatrix berechnet:

```
(1/4  1/4  1/4  1/4)
(  1  -1   0   0)
(  0  -1   1   0)
(  0  -1   0   1)
```

## Helmert

**Helmert-Kontraste.** Hierbei werden die Kategorien einer unabhängigen Variablen mit dem Mittelwert der nachfolgenden Kategorien verglichen. Die allgemeine Matrixform lautet:

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

Hierbei entspricht  $k$  der Anzahl der Kategorien in der unabhängigen Variablen. Eine unabhängige Variable mit vier Kategorien weist beispielsweise eine Helmert-Kontrastmatrix der folgenden Form auf:

```
(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)
```

## Differenz

**Differenzkontraste oder umgekehrte Helmert-Kontraste.** Hierbei werden die Kategorien einer unabhängigen Variablen mit dem Mittelwert der vorausgehenden Kategorien der Variablen verglichen. Die allgemeine Matrixform lautet:

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

Hierbei entspricht  $k$  der Anzahl der Kategorien in der unabhängigen Variablen. Die Differenzkontraste für eine unabhängige Variable mit vier Kategorien lauten beispielsweise wie folgt:

```
( 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)
```

## Polynomial

**Orthogonale polynomiale Kontraste.** Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien, der zweite Freiheitsgrad den quadratischen Effekt, der dritte Freiheitsgrad den kubischen Effekt und so weiter für Effekte höherer Ordnung.

Sie können den Abstand zwischen den Stufen der von der angegebenen kategorialen Variable gemessenen Behandlung angeben. Gleiche Abstände (die Standardeinstellung, wenn keine Metrik angegeben wird) können als aufeinander folgende Ganzzahlen von 1 bis  $k$  angegeben werden, wobei  $k$  der Anzahl der Kategorien entspricht. Wenn die Variable *Wirkstf* drei Kategorien aufweist, entspricht der Unterbefehl

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

entspricht dem

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

Gleiche Abstände sind jedoch nicht immer erforderlich. Angenommen, *Wirkstf* stellt verschiedene Dosierungen eines Wirkstoffs dar, der drei verschiedenen Gruppen verabreicht wurde. Wenn der zweiten Gruppe eine doppelt so hohe Dosierung wie der ersten Gruppe und der dritten Gruppe eine dreimal so



hohe Dosierung wie der ersten Gruppe verabreicht wurde, weisen die Behandlungskategorien gleiche Abstände auf. Hierbei eignet sich die folgende, aus aufeinander folgenden Ganzzahlen bestehende Metrik:

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

Wenn der zweiten Gruppe jedoch eine viermal so hohe Dosierung wie der ersten Gruppe und der dritten Gruppe eine siebenmal so hohe Dosierung wie der ersten Gruppe verabreicht wurde, eignet sich die folgende Metrik:

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

In beiden Fällen bedeutet diese Angabe des Kontrastes, dass der erste Freiheitsgrad für *Wirkstf* den linearen Effekt und der zweite Freiheitsgrad den quadratischen Effekt der Dosierung enthält.

Polynomiale Kontraste sind insbesondere bei Tests auf Trends und bei der Untersuchung von Wirkungsf lächen nützlich. Polynomiale Kontraste können auch für die nicht lineare Kurvenanpassung verwendet werden, beispielsweise für kurvilineare Regression.

## Wiederholt

**Vergleich von aufeinander folgenden Stufen einer unabhängigen Variablen.** Die allgemeine Matrixform lautet:

```
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)
```

Hierbei entspricht  $k$  der Anzahl der Kategorien in der unabhängigen Variablen. Die wiederholten Kontraste für eine unabhängige Variable mit vier Kategorien lauten beispielsweise wie folgt:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Diese Kontraste sind bei der Profilanalyse und in Situationen nützlich, in denen Differenzwerte benötigt werden.

## Spezial

**Benutzerdefinierter Kontrast.** Hierbei können Sie spezielle Kontraste in Form einer quadratischen Matrix angeben, wobei die Anzahl der Zeilen und Spalten der Anzahl der Kategorien in der unabhängigen Variablen entsprechen muss. Bei MANOVA und LOGLINEAR ist die erste Zeile stets der Mittelwerteffekt (konstanter Effekt). Sie stellt das Set der Gewichtungen dar, mit denen angegeben wird, wie die Mittelwerte anderer unabhängiger Variablen (sofern vorhanden) über die vorliegende Variable ermittelt werden. Im Allgemeinen ist dieser Kontrast ein Vektor, der aus Einsen besteht.

Die verbleibenden Zeilen der Matrix enthalten die speziellen Kontraste, mit denen die Vergleiche zwischen den Kategorien der Variable angegeben werden. In der Regel sind orthogonale Kontraste am nützlichsten. Orthogonale Kontraste sind statistisch unabhängig und frei von Redundanz. Kontraste sind orthogonal, wenn die folgenden Bedingungen zutreffen:

- Die Summe der Kontrastkoeffizienten in jeder Zeile beträgt 0.
- Die Summe der Produkte der entsprechenden Koeffizienten aller Paare in disjunkten Zeilen beträgt ebenfalls 0.

Angenommen, die Behandlung weist vier Stufen auf, und Sie möchten die verschiedenen Behandlungsstufen miteinander vergleichen. Hierfür eignet sich der folgende spezielle Kontrast:

```
(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
```

Bei MANOVA, LOGISTIC REGRESSION und COXREG verwenden Sie zur Angabe dieses Kontrasts den folgenden Unterbefehl vom Typ CONTRAST:

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

Bei LOGLINEAR müssen Sie Folgendes angeben:

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

Jede Zeile mit Ausnahme der in der Zeile angegebenen Zeilensummen beträgt 0. Die Summe der Produkte aller Paare in disjunkten Zeilen beträgt ebenfalls 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
```

Spezielle Kontraste müssen nicht notwendigerweise orthogonal sein. Sie dürfen jedoch keine linearen Kombinationen voneinander darstellen. Falls dies doch der Fall ist, meldet die Prozedur die lineare Abhängigkeit und die Verarbeitung wird abgebrochen. Helmert-, Differenz- und polynomiale Kontraste sind orthogonal.

## Indikator

**Codierung für Indikatorvariablen.** Auch bekannt als Dummy-Codierung. Dieser Kontrast ist bei LOGLINEAR und MANOVA nicht verfügbar. Die Anzahl der neu codierten Variablen entspricht  $k-1$ . Fälle in der Referenzkategorie werden für alle  $k-1$  Variablen als 0 codiert. Ein Fall in der Kategorie  $I^{\text{th}}$  ist für alle Indikatorvariablen mit Ausnahme des  $I^{\text{th}}$ , das 1 codiert ist, mit 0 codiert.

## Kernel-Ridge-Regression

Kernel Ridge Regression ist eine Erweiterungsprozedur, die die Python-Klasse **sklearn.kernel\_ridge.KernelRidge** verwendet, um die Kernel-Ridge-Regressionsmodelle zu schätzen. Kernel-Ridge-Regressionsmodelle sind nichtparametrische Regressionsmodelle, die in der Lage sind, lineare und nichtlineare Beziehungen zwischen Prädiktorvariablen und -ergebnissen zu modellieren. Die Ergebnisse können hochsensitiv für die Auswahl von Modellhyperparametern sein. Die Kernel-Ridge-Regression erleichtert die Auswahl von Hyperparameterwerten durch die k-fache Kreuzvalidierung auf spezifizierten Rastern von Werten, die die Klasse **sklearn.model\_selection.GridSearchCV** verwenden.

### Beispiel

#### Statistik

Additive\_CHI2, CHI2, Kosinus, Laplace, Linear, Polynomial, RBF, Sigma, Alpha, Gamma, Coef0, Grad, Kreuzvalidierung, beobachtet versus vorhergesagt, Residuen, versus vorhergesagte, duale Gewichts-Koeffizienten, Kernel-Raum-Gewichts-Koeffizienten.

## Erläuterung der Daten

### Daten

- Sie können eine beliebige oder alle der acht verschiedenen Kernelfunktionen angeben.
- Die ausgewählte Kernelfunktion bestimmt, welche Hyperparameter aktiv sind.
- Zu den Hyperparametern gehören Alpha für die Ridge-Regularisierung, die allen Kernen gemeinsam sind, plus drei weitere Hyperparameter für jede spezifische Kernel-Funktion.
- Wenn mehrere Kernel-Unterbefehle angegeben werden oder mehr als ein Wert für einen beliebigen Parameter angegeben wird, wird eine Rastersuche mit der Quervalidierung für die Auswertung von Modellen durchgeführt und das beste passende Modell, das auf "ferngehaltenen" Daten basiert, ausgewählt.

- Die Erweiterung nimmt getrennte Variablen aus der Prozedur "Datei aufteilen" an und gewichtet diese anhand der Prozedur "Fälle gewichten".
- Wenn Gewichtungen eingeschlossen werden, werden sie bei der Erstellung von angepassten Werten in allen Analysen verwendet. Aufgrund von Einschränkungen in der Bewertungsmethode in der Klasse `sklearn.model_selection.GridSearchCV` werden Kreuzvalidierungsauswertungen, die für die Modellauswahl verwendet werden, nicht gewichtet.

## Voraussetzungen

### Abrufen einer Kernel-Ridge-Regression

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Regression > Kernel Ridge...**

2. Wählen Sie eine **Abhängige** Variable aus.

3. Wählen Sie eine oder mehrere **Unabhängige** Variablen aus.

4. Die Standardeinstellung **Einzelnes Modell** wird verwendet, wenn nur ein Wert für jeden Kernelfunktionsparameter angegeben wird. Wenn die Einstellung **Einzelnes Modell** ausgewählt ist, können Sie keine zusätzlichen **Kernel**-Funktionen angeben, und die Gewichtungen werden während der gesamten Analyse, der Auswertung und der Bewertung von Ergebnissen vollständig angewendet. Sie können auch die Aufwärts- und Abwärtspfeilsteuerelemente verwenden, um die Kernelfunktionen neu zu ordnen.

Wählen Sie optional **Modellauswahl** in der Liste **Modus** aus.

Wenn **Modellauswahl** aus der Liste **Modus** ausgewählt wird, können Sie der **Kernel**-Liste mehrere Kernelfunktionen hinzufügen.

- Klicken Sie auf das Steuerelement "Hinzufügen" (+), um zusätzliche Kernelfunktionen aufzunehmen.
- Klicken Sie auf die leere Zelle in der Spalte **Kernel**, um eine Kernelfunktion auszuwählen.
- Klicken Sie doppelt auf eine beliebige Kernelfunktionszeilenzelle, um die Werte für die Kernelfunktionsparameter für die entsprechende Spalte (**Alpha**, **Gamma**, **Coef0**, **Grad**) anzugeben. Weitere Informationen finden Sie unter „Kernelparameter“ auf Seite 32. Die Standardparameter für die Optimierung der Kernelfunktion sind im Folgenden aufgelistet.

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

#### **Kosinus**

ALPHA=1

#### **Laplace**

ALPHA=1 GAMMA=1/p

#### **Linear**

Die Standard-Kernelfunktion. ALPHA=1

#### **Polynomial**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

#### **RBF**

ALPHA=1 GAMMA=1/p

#### **Sigma**

ALPHA=1 GAMMA=1/p COEF0=1

**Anmerkung:** Wenn mehr als ein Wert für einen Kernelfunktionsparameter angegeben wird, wird eine Rastersuche mit der Quervalidierung für die Auswertung von Modellen durchgeführt und das beste Anpassungsmodell, das auf der Grundlage der ausgehaltenen Daten basiert, ausgewählt.

5. Klicken Sie optional auf **Optionen** , um die Anzahl der Kreuzvalidierungsfalten, Anzeigeoptionen, Ploteinstellungen und zu speichernde Elemente anzugeben. Weitere Informationen finden Sie unter „Kernel-Ridge-Regression: Optionen“ auf Seite 32.
6. Klicken Sie auf **OK**.

## Kernelparameter

Der Dialog **Kernelparameter** stellt Optionen für die Angabe einzelner Kernelfunktionsparameterwerte und für die Angabe, dass die Modellauswahl mithilfe einer Rastersuche über die Kombinationen von Kernen und angegebenen Rasterparameterwerten durchgeführt wird, zur Verfügung.

### Einzelne Parameter angeben

Aktivieren Sie die Einstellung, um Werte für den ausgewählten Kernelfunktionsparameter anzugeben.

- Geben Sie einen Wert ein, und klicken Sie auf **Hinzufügen**, um den Wert in den Kernelfunktionsparameter aufzunehmen.
- Wählen Sie einen Parameterwert aus und klicken Sie auf **Ändern**, um den Wert zu aktualisieren.
- Wählen Sie einen Parameterwert aus und klicken Sie auf **Entfernen** , um den Wert zu löschen.

### Rasterparameter angeben

Aktivieren Sie die Einstellung, um anzugeben, dass die Modellauswahl mithilfe einer Rastersuche über die Kombinationen von Kernen und den angegebenen Rasterparameterwerten ausgeführt wird.

## Kernel-Ridge-Regression: Optionen

Das Dialogfeld **Diagramme** bietet Optionen für die Angabe der Anzahl von Kreuzvalidierungsfalten, Anzeigeoptionen, Ploteinstellungen und zu speichernden Elementen.

### Anzahl der Kreuzvalidierungsfalten

Die Anzahl der Splits oder Falten in der Kreuzvalidierung mit Rastersuche zur Modellauswahl. Geben Sie einen ganzzahligen Wert ein, der größer ist als 1. Der Standardwert ist 5. Die Einstellung ist nur verfügbar, wenn **Modellauswahl** im primären **Kernel-Ridge-Regression**-Dialogfeld als **Modus** ausgewählt wird.

### Anzeige

Bietet Optionen für die Angabe, welche Ausgabe angezeigt werden soll, wenn die Kreuzvalidierung wirksam ist.

#### Empfohlen

In der Standardeinstellung werden nur grundlegende Ergebnisse für das ausgewählte beste Modell angezeigt.

#### Vergleichen

Zeigt grundlegende Ergebnisse für alle ausgewerteten Modelle an.

#### Modelle und Falten vergleichen

Zeigt für jedes ausgewertete Modell vollständige Ergebnisse für jede Aufteilung oder Faltung an.

### Diagramm

Stellt Optionen für die Angabe von Werten für beobachtete oder Restwerte im Vergleich zu vorhergesagten Werten bereit.

#### Beobachtet vs. vorhergesagt

Zeigt ein Streudiagramm der beobachteten und vorhergesagten Werte für das angegebene oder beste Modell an.

#### Residuen vs. Vorhergesagt

Zeigt ein Streudiagramm der Residuen gegenüber den vorhergesagten Werten für das angegebene oder beste Modell an.

### Speichern

Die Tabelle enthält Optionen zur Angabe von Variablen, die in dem aktiven Dataset gespeichert werden sollen.

### **Vorhergesagte Werte**

Speichert vorhergesagte Werte aus dem angegebenen oder besten Modell in das aktive Dataset. Ein optionaler Variablenname kann aufgenommen werden.

### **Residuen**

Speichert Residuen aus den angegebenen oder besten Modellvorhersagen in das aktive Dataset. Ein optionaler Variablenname kann aufgenommen werden.

### **Duale Koeffizienten**

Speichert Dual-oder Kernel-Raum-Gewichts-Koeffizienten vom angegebenen Modell in den aktiven Datensatz. Ein optionaler Variablenname kann aufgenommen werden. Die Einstellung ist nicht verfügbar, wenn **Modellauswahl** im primären **Kernel-Ridge-Regression** -Dialogfeld als **Modus** ausgewählt wird.

## **Parametrische beschleunigte Fehlerzeitmodelle**

---

Eine AFT-Modellanalyse (AFT = Parametric Accelerated Failure Time) ruft die Prozedur für parametrische Überlebensmodelle mit nicht rekursiven Lebenszeitdaten auf. Parametrische Überlebensmodelle gehen davon aus, dass die Überlebenszeit einer bekannten Verteilung folgt, und diese Analyse passt Modelle für beschleunigte Ausfallzeiten mit ihren Modelleffekten proportional zur Überlebenszeit an.

### **Abrufen einer Analyse für parametrische Modelle für beschleunigte Fehlerzeiten**

1. Wählen Sie in den Menüs Folgendes aus:

**Analyse > Überleben > Parametrische AFT-Modelle (AFT-Accelerated Failure Time)**

2. Wählen Sie eine Quellenvariable aus.

#### **Zeit**

##### **Überleben**

Einzelne numerische Variable, die die Dauer der Überlebenszeit angibt.

##### **Start/Ende**

Numerische Variablen, die **Startzeit** und **Endzeit** angeben.

#### **Status**

Einzelne optionale Zeichenfolge oder numerische Variable, die eine der folgenden Stauseinstellungen bestimmt:

##### **Fehler/Ereignis**

Ordnet einen Datensatz einer Fehler-/Ereigniskategorie zu. Der Standardwert für eine Zeichenfolgestatusvariable ist F.

##### **Rechtszensierung**

Ordnet einen Datensatz einer rechtszensierenden Kategorie zu. Der Standardwert für eine Zeichenfolgestatusvariable ist R.

##### **Linkszensierung**

Ordnet einen Datensatz einer linkszensierenden Kategorie zu. Der Standardwert für eine Zeichenfolgestatusvariable ist L.

##### **Intervallzensierung**

Ordnet einen Datensatz einer Intervallzensierungskategorie zu. Nur für **Start/Ende**. Der Standardwert für eine Zeichenfolgestatusvariable ist I.

##### **Verfahren für nicht zugeordnete Werte**

Steuert, welcher Kategorie die nicht zugeordneten Datensätze zugeordnet werden sollen. Um die Datensätze zu löschen, deren Zuordnung fehlgeschlagen ist, wählen Sie **Aus Analyse ausschließen**.

Für **Überleben** lautet der Standardstatus für alle Fälle **Fehler/Ereignis**. Für **Start/Ende** lautet der Standardstatus **Interval Censoring**. Klicken Sie auf die Schaltfläche **Ereignis definieren**, um ein Ereignis für die Statusvariable zu definieren.

**Kovariate(n)**

Eine oder mehrere optionale numerische Variablen, die als Kovariaten behandelt werden sollen. Beachten Sie, dass eine Variable nicht sowohl durch **Kovariaten** als auch durch **feste Faktoren** angegeben werden kann.

**Feste Faktoren**

Eine oder mehrere optionale Variablen, die als Faktoren behandelt werden sollen. Eine Variable kann nicht sowohl durch **feste Faktoren** als auch durch **Kovariaten** angegeben werden.

**Linkstrunkierung**

Einzelne optionale numerische Variable für linkes Abschneiden nur für **Survival**.

## Parametrische beschleunigte Fehlerzeitmodelle: Kriterien

**Kriterien**

Eine optionale Anzeige zur Angabe der allgemeinen Kriterien.

**Konfidenzintervall**

Ein optionaler Prozentsatz zur Angabe des Niveaus für die Konfidenzintervalle von Regressionsparametern. Es muss ein einzelner numerischer Wert zwischen 0 und 100 sein. Der Standardwert ist 95.

**Fehlende Werte**

Eine Option zum Steuern, wie die benutzerdefiniert fehlenden Werte behandelt werden:

**Benutzer- und systemdefiniert fehlende Werte ausschließen**

Behandelt die benutzerdefiniert fehlenden Werte als gültige Werte. Dies ist die Standardeinstellung.

**Benutzerdefiniert fehlende Werte sind gültig**

Ignoriert die benutzerdefiniert fehlenden Wertbezeichnungen und behandelt sie als gültige Werte.

**Statusbehandlung**

Nur für **Start/Ende**. Eine Option, mit der gesteuert werden kann, wie Datensätze mit falschen Statusfeldern behandelt werden:

**Datensatz mit Konflikten verwerfen**

Löscht die Datensätze mit Konflikten. Dies ist die Standardeinstellung.

**Abrufen der Zeitinformationen nach Status**

Ruft die Zeitinformationen entsprechend dem Status ab

**Status anhand der Zeitinformationen ableiten**

Ändert den Status entsprechend den Zeitinformationen.

## Modelle für parametrische beschleunigte Ausfallzeit: Modell

**Modell**

Eine optionale Anzeige zur Angabe der Modelloptionen und -einstellungen.

**Verteilung der Überlebenszeit**

Eine Option zur Angabe der Verteilung der Überlebenszeit.

**Weibull**

Gibt die Weibull-Verteilung an. Dies ist die Standardeinstellung.

**Exponentiell**

Gibt die Exponentialverteilung an

**Logarithmische Normalverteilung**

Gibt die Log-Normal-Verteilung an

**Log-logistisch**

Gibt die loglogistische Verteilung an.

### **Kovariateneinstellungen**

Kovariatenvariablen angeben.

### **Faktoreinstellungen**

Geben Sie Faktorvariablen an.

### **Anfangswert des konstanten Terms**

Eine Option zur Angabe des Anfangswerts des konstanten Terms. Wenn er angegeben wird, muss er ein einzelner numerischer Wert sein und darf nicht 0 sein.

### **Anfangswert des Skalenparameters**

Eine Option zur Steuerung der Einstellung des Skalenparameters.

### **Standardfehler der entsprechenden OLS-Regression**

Verwendet den Standardfehler der entsprechenden gewöhnlichen Regression der kleinsten Quadrate als Anfangswert.

### **Invertiert den Standardfehler der entsprechenden OLS-Regression**

Verwendet den reziproken Wert des Standardfehlers.

### **Vom Benutzer angegebener Wert**

Wenn ein einzelner numerischer Wert angegeben wird, wird der Wert als Anfangswert verwendet. Falls angegeben, muss er größer als 0 sein.

## **Parametrische beschleunigte Fehlerzeitmodelle: Schätzung**

### **Schätzung**

Eine optionale Anzeige zur Angabe der Einstellungen zur Steuerung der Schätzung der beschleunigten Fehlerzeitmodelle und des Auswahlprozesses für optionale Features.

### **Methode mit alternierender Richtung oder Multiplikatoren (ADMM)**

#### **Schnell**

Wendet die schnelle Wechselrichtungsmethode von Multiplikatoren (ADMM) an. Dies ist die Standardeinstellung.

#### **Traditionell**

Wendet den traditionellen ADMM-Algorithmus an.

### **L-1-Regularisierung anwenden**

Führt den Prozess zur Steuerung der Merkmalauswahl durch Das Feld **Penalty-Parameter** gibt den Penalisiertungsparameter an, der den Regularisierungsprozess steuert. Der Wert muss ein Einzelwert größer als 0 sein. Die Standardeinstellung ist 0.001.

### **Modellkonvergenzkriterien**

#### **Parameterkonvergenz**

Gibt die Konvergenzkriterien für den Parameter an Es muss ein einzelner numerischer Wert sein, der zu  $[0, 1)$  gehört. Die Standardeinstellung ist 0,000001. Als **Typ** können Sie entweder **ABSOLUTE** auswählen, um die absolute Konvergenz auf die innere Optimierung anzuwenden, oder **RELATIVE**, um die relative Konvergenz auf die innere Optimierung anzuwenden. Das optionale Feld **Wert** gibt ein Schlüsselwort an.

#### **Konvergenz der Zielfunktion**

Gibt die Konvergenzkriterien für die Zielfunktion an. Es muss ein einzelner numerischer Wert sein, der zu  $[0, 1)$  gehört. Die Standardeinstellung ist 0, d. h., die Konvergenzkriterien werden nicht angewendet. Als **Typ** können Sie entweder **ABSOLUTE** auswählen, um die absolute Konvergenz auf die innere Optimierung anzuwenden, oder **RELATIVE**, um die relative Konvergenz auf die innere Optimierung anzuwenden. Das optionale Feld **Wert** gibt ein Schlüsselwort an.

#### **Konvergenz der Hesse-Matrix**

Gibt die Konvergenzkriterien für die Hessische Matrix an. Es muss ein einzelner numerischer Wert sein, der zu  $[0, 1)$  gehört. Die Standardeinstellung ist 0, d. h., die Konvergenzkriterien werden nicht angewendet. Als **Typ** können Sie entweder **ABSOLUTE** auswählen, um die absolute Konver-

genz auf die innere Optimierung anzuwenden, oder **RELATIVE**, um die relative Konvergenz auf die innere Optimierung anzuwenden. Das optionale Feld **Wert** gibt ein Schlüsselwort an.

### **Konvergenzkriterien für Residuen**

Eine Option zur Steuerung des Optimierungsprozesses.

#### **Sowohl primales als auch duales Residuum**

Wendet sowohl das ursprüngliche als auch das doppelte Konvergenzkriterium an. Dies ist die Standardeinstellung.

#### **Nur primales Residuum**

Wendet das Konvergenzkriterium der ursprünglichen Residuen an.

#### **Nur duales Residuum**

Wendet das doppelte Konvergenzkriterium für Residuen an

### **Methode**

Ein optionaler Parameter zur Angabe der Schätzmethode.

#### **Automatisch**

Wählt die Methode automatisch basierend auf dem Stichprobendataset aus. Dies ist die Standardeinstellung. Das Feld **Schwellenwertanzahl der Prädiktoren** gibt den Schwellenwert für die Anzahl der Prädiktoren an und muss eine einzelne ganze Zahl größer als 1 sein. Der Standardwert beträgt "1000".

#### **Newton-Raphson**

Wendet die Newton-Raphson-Methode an.

#### **L-BFGS**

Wendet den BFGS-Algorithmus mit begrenztem Speicher an. Das Feld **Update** gibt die Anzahl der früheren Aktualisierungen an, die vom BFGS-Algorithmus mit begrenztem Speicher verwaltet werden, und muss eine einzelne ganze Zahl größer-gleich 1 sein. Der Standardwert ist 5.

### **Iteration**

#### **Maximalzahl der Iterationen**

Gibt die maximale Anzahl der Iterationen an. Es muss eine einzelne ganze Zahl sein, die zu [1, 100] gehört. Die Standardeinstellung ist 20.

#### **Maximale Schritthalbierung**

Gibt die maximale Anzahl der Schritthalbierung an. Es muss eine einzelne ganze Zahl sein, die zu [1, 20] gehört. Die Standardeinstellung ist 5.

#### **Maximale Anzahl von Zeilensuchen**

Gibt die maximale Anzahl der Zeilensuchvorgänge an. Es muss eine einzelne ganze Zahl sein, die zu [1, 100] gehört. Die Standardeinstellung ist 20.

#### **Absolute Konvergenz für Iterationsprozess**

Gibt die absolute Konvergenz für den äußeren Iterationsprozess an. Es muss ein einzelner numerischer Wert sein, der zu (0, 1) gehört. Die Standardeinstellung ist 0.0001.

#### **Relative Konvergenz für Iterationsprozess**

Gibt die relative Konvergenz für den äußeren Iterationsprozess an. Es muss ein einzelner numerischer Wert sein, der zu (0, 1) gehört. Die Standardeinstellung ist 0.01.

## **Parametrische beschleunigte Fehlerzeitmodelle: Drucken**

### **Drucken**

Eine optionale Anzeige zur Steuerung der Tabellenausgaben.

#### **Details zur Faktorcodierung**

Bei Auswahl dieser Option werden die Codierungsdetails der Faktoren angezeigt und gedruckt. Der Prozess wird ignoriert, wenn keine Faktoren wirksam sind.



### **Den Regressionsparametern zugeordnete Anfangswerte**

Wenn diese Option ausgewählt ist, werden die im Schätzprozess verwendeten Anfangswerte angezeigt.

### **Verlauf der Modelliteration**

Bei Auswahl dieser Option wird der Iterationsverlauf der Überlebensanalyse angezeigt. Geben Sie im Feld **Anzahl Schritte** die Anzahl der Schritte zwischen 1 und 99999999 an. Die Standardeinstellung ist 1.

### **Auswahlergebnisse mit**

Steuert die Anzeige der Details der Featureauswahl.

#### **Sowohl ausgewählte als auch nicht ausgewählte Variablen**

Sowohl ausgewählte als auch nicht ausgewählte Variablen in der Tabelle anzeigen.

#### **Nur ausgewählte Variablen**

Nur die ausgewählten Variablen anzeigen.

#### **Nur nicht ausgewählte Variablen**

Nur die nicht ausgewählten Variablen anzeigen. Das Feld **Maximale Anzahl anzuzeigender Variablen** gibt die maximale Anzahl der in der Tabelle ausgegebenen Variablen an. Die Standardeinstellung ist 30.

## **Parametrische Modelle für beschleunigte Ausfallzeiten: Vorhersagen**

### **Vorhersage**

Eine optionale Anzeige zum Bewerten und Speichern der vorhergesagten Statistiken im aktiven Dataset.

### **Zeitwerte für Scoring**

#### **Durch abhängige Variable(n) definierte Zeitwerte**

Bewertet die **Vorhersagen** basierend auf der Zeitvariablen, die für das parametrische Überlebensmodell angegeben wurde

#### **Regelmäßige Intervalle**

Bewertet die **Vorhersagen** basierend auf zukünftigen Zeitwerten. Das Feld **Zeitintervall** gibt das Zeitintervall an und muss ein einzelner numerischer Wert größer als 0 sein. Das Feld **Anzahl Zeiträume** gibt die Anzahl der Zeiträume an und muss eine einzelne numerische Ganzzahl zwischen 2 und 100 sein.

#### **Zeitdifferenz**

Bewertet die **Vorhersagen** basierend auf der Zeitdauer, um die zukünftigen Zeitwerte zu definieren. Es muss sich um eine einzelne numerische Variable handeln.

### **Vorhersagen**

#### **Überleben**

Bewertet und speichert die vorhergesagten Überlebensstatistiken im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredSurvival`.

#### **Gefahr**

Bewertet und speichert die vorhergesagten Gefahren im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredHazard`.

#### **Kumulierte Gefahr**

Bewertet und speichert die vorhergesagten kumulativen Gefahren im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredCumHazard`.

#### **Bedingtes Überleben**

Bewertet und speichert die vorhergesagten bedingten Überlebensstatistiken im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredConditionalSurvival`. Der Prozess wird ignoriert, wenn `PASTTIME` nicht angegeben ist. Ein Wert für **Vergangene Überlebenszeit** ist erforderlich und gibt die vergangenen Zeitwerte für das Scoring an. Es muss sich um eine einzelne numerische Variable handeln.

## Modelle für parametrische beschleunigte Ausfallzeit: Diagramm

### Diagramm

#### Diagramm zu Cox-Snell-Residuen

Wählen Sie **Diagramm anzeigen** aus, um ein Cox-Snell-Residuendiagramm zu erstellen. Geben Sie im Feld **Anzahl Klassiertrennpunkte** eine Zahl von 1 bis 10000 an. Die Standardeinstellung ist 100.

#### Funktionsdiagramme

Eine Option zur Steuerung der Funktionsdiagramme.

#### Typ

##### Überleben

Erstellt das Diagramm für Überlebensfunktionen.

##### Gefahr

Erstellt das Diagramm für die Gefahrenfunktionen.

##### Dichte

Erstellt ein Diagramm für die Dichtefunktionen.

#### Anzahl anzuzeigender Punkte

Gibt die Anzahl der Funktionspunkte zwischen 1 und 200 an. Die Standardeinstellung ist 100.

#### Kovariatenwerte für Diagramm

Eine optionale Angabe der vom Benutzer bereitgestellten Werte und Zuweisung zu den Prädiktoren. Standardmäßig werden die designierten Diagramme mit dem **Mittelwert** jeder wirksamen Kovariate und der Kategorienhäufigkeit jedes Faktors erstellt. Bei Angabe werden die designierten Diagramme basierend auf der Einstellung des Musters erstellt. Wenn doppelte Variablen vorhanden sind, wird die zuerst angegebene Variable erkannt und der Rest ignoriert. Ein Modelleffekt muss eine gültige Variable enthalten. Bei einer Kovariate muss der benutzerdefinierte Wert numerisch sein. Wenn eine Variable ausgelassen wird, bedeutet dies, dass die Kategorienhäufigkeit und der **Mittelwert** standardmäßig für den Faktor bzw. die Kovariate verwendet werden. Wenn einer Variablen ein ungültiger Wert zugeordnet wird, wird das angeforderte Muster nicht dargestellt.

#### Faktorwerte für Diagramm

Eine optionale Angabe der vom Benutzer bereitgestellten Werte und Zuweisung zu den Prädiktoren. Wenn doppelte Variablen vorhanden sind, wird die zuerst angegebene Variable erkannt und der Rest ignoriert. Ein Modelleffekt muss eine gültige Variable enthalten. Die Auslassung einer Variablen im Effekt gibt an, dass die Kategorienhäufigkeit und der Mittelwert standardmäßig für den Faktor bzw. die Kovariate verwendet werden. Wenn einer Variablen ein ungültiger Wert zugeordnet wird, wird das angeforderte Muster nicht dargestellt.

#### Separate Zeilen für

Eine Option zur Angabe einer kategorialen Variablen, mit der die Liniendiagramme gezeichnet werden sollen.

#### Maximale Anzahl von Linien in einem Diagramm

Gibt die maximale Anzahl der Linien in einem Diagramm an, wenn **Separate Linien für** angegeben ist. Die Standardeinstellung ist 10.

## Parametrische beschleunigte Fehlerzeitmodelle: Exportieren

### Exportieren

Wählen Sie **Modellinformationen in XML-Datei exportieren** aus, um die Modell- und Parameterinformationen zum Scoring in eine PMML-Datei zu schreiben. Sie müssen das Verzeichnis und den Dateinamen der PMML-Datei angeben, die gespeichert werden soll.

## Survival AFT Ereignisse für Statusvariablen definieren

Vorkommen des ausgewählten Werts bzw. der ausgewählten Werte für die Statusvariable geben an, dass das terminale Ereignis für diese Fälle aufgetreten ist. Alle anderen Fälle werden als zu zensieren betrachtet. Geben Sie einen einzelnen Wert oder einen Bereich von Werten ein, die das zu untersuchende Ereignis kennzeichnen.

## Parametrische beschleunigte Fehlerzeitmodelle: Kategorie auswählen

Die Einstellung "Kategorie auswählen" bietet Optionen zur Auswahl eines Werts, der die zu modellierende Warengruppe als Vergleichsbasis bezeichnet.

Kategorie auswählen

Klicken Sie auf 'Letzte Kategorie', um das Dialogfenster 'Kategorie auswählen' aufzurufen.

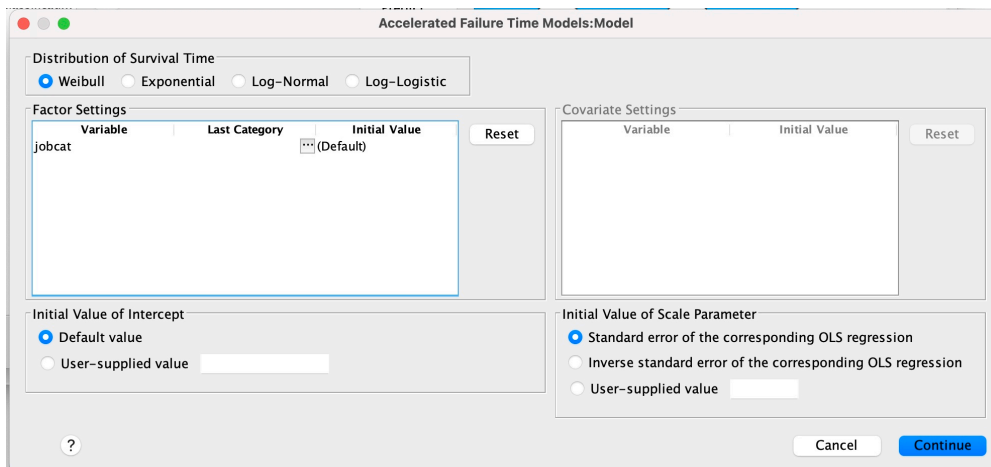


Abbildung 1. Modelle für beschleunigte Lebensdauer-Dialogfenster-Kategorie

Um eine Kategorie als Baseline festzulegen, wählen Sie im Dialogfeld "Kategorie auswählen" einen Wert aus.

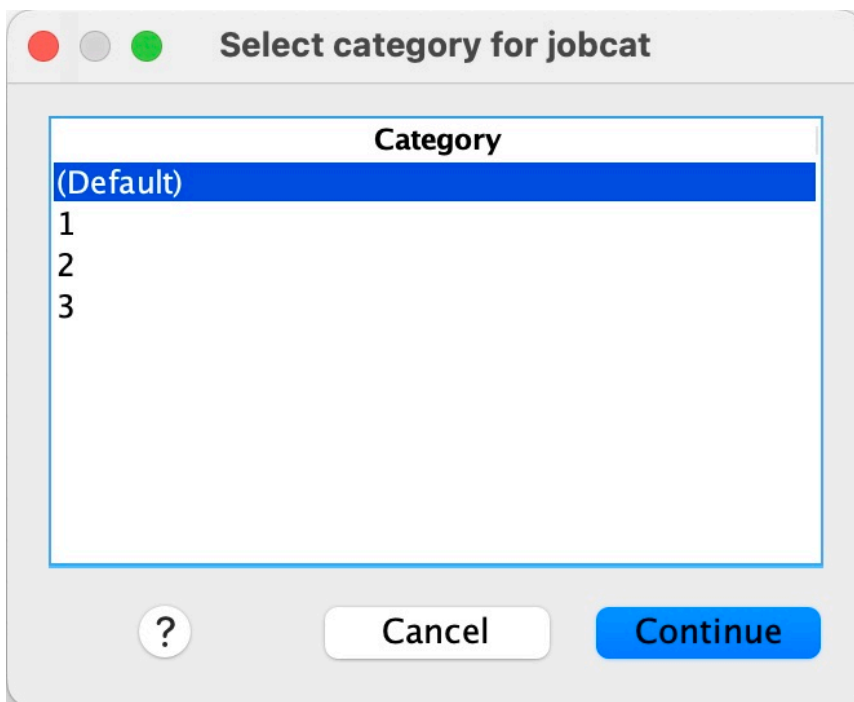


Abbildung 2. Modelle für beschleunigte Lebensdauer-Dialogfenster-Kategorie auswählen

Klicken Sie auf Weiter.

## Parametrische gemeinsam genutzte Frailty-Modelle

---

Eine parametrische Überlebensanalyse für gemeinsam genutzte Frailty-Modelle startet die Prozedur für parametrische Überlebensmodelle mit der Eingabe von Daten für die rekursive Lebensdauer. Parametrische Überlebensmodelle gehen davon aus, dass die Überlebenszeit einer bekannten Verteilung folgt, und diese Analyse integriert einen Frailty-Term in ein parametrisches Überlebensmodell. Er wird als Zufallskomponente behandelt, um einen nicht beobachteten Effekt aufgrund der Variabilität auf Individual- oder Gruppenebene zu berücksichtigen.

### Abrufen einer Analyse für parametrische gemeinsam genutzte Frailty-Modelle

1. Wählen Sie im Menü Folgendes aus:

**Analysieren > Überleben > Parametrische gemeinsam genutzte Frailty-Modelle**

2. Wählen Sie eine Quellenvariable aus.

#### Zeit

##### Überleben

Die Überlebenszeit wird durch eine Variable dargestellt, um die Endzeit anzugeben. Die Startzeit wird auf 0 gesetzt.

##### Start/Ende

Numerische Variablen, die die **Startzeit** und die **Endzeit** angeben.

#### Thema

Erforderlich, um die Prozedur auszuführen. Gibt eine einzelne Variable für die Subjekt-ID an

#### Intervallskala

Gibt eine einzelne numerische Variable für die Intervallnummer an, die verwendet wird, um die verschiedenen wiederkehrenden Datensätze zu identifizieren, die dieselbe Subjekt-ID verwenden.

#### Status

Einzelne optionale Zeichenfolge oder numerische Variable, die eine der folgenden Stauseinstellungen bestimmt:

##### Fehler/Ereignis

Ordnet einen Datensatz einer Fehler-/Ereigniskategorie zu. Der Standardwert für eine Zeichenfolgestatusvariable ist F.

##### Rechtszensurierung

Ordnet einen Datensatz einer rechtszensierenden Kategorie zu. Der Standardwert für eine Zeichenfolgestatusvariable ist R.

##### Verfahren für nicht zugeordnete Werte

Steuert, welcher Kategorie die nicht zugeordneten Datensätze zugeordnet werden sollen. Um die Datensätze zu löschen, deren Zuordnung fehlgeschlagen ist, wählen Sie **Aus Analyse ausschließen**.

Klicken Sie auf die Schaltfläche **Ereignis definieren**, um ein Ereignis für die Statusvariable zu definieren.

#### Kovariate(n)

Eine oder mehrere optionale numerische Variablen, die als Kovariaten behandelt werden sollen. Beachten Sie, dass eine Variable nicht sowohl durch **Kovariate (n)** als auch durch **feste Faktoren** angegeben werden kann.

#### Feste Faktoren

Eine oder mehrere optionale Variablen, die als Faktoren behandelt werden sollen. Eine Variable kann nicht sowohl durch **feste Faktoren** als auch durch **Kovariaten** angegeben werden.

## Parametrische gemeinsam genutzte Frailty-Modelle: Kriterien

### Kriterien

Eine optionale Anzeige zur Angabe der allgemeinen Kriterien.

#### Konfidenzintervall

Ein optionaler Prozentsatz zur Angabe des Niveaus für die Konfidenzintervalle von Regressionsparametern. Es muss ein einzelner numerischer Wert zwischen 0 und 100 sein. Der Standardwert ist 95.

#### Signifikanzniveau

Eine Option zur Angabe des Signifikanzniveaus des Likelihood-Quotienten-Tests für die Fragmentierungskomponente. Es muss ein einzelner numerischer Wert zwischen 0 und 1 sein. Die Standardeinstellung ist 0.05.

#### Fehlende Werte

Eine Option zum Steuern, wie die benutzerdefiniert fehlenden Werte behandelt werden:

##### Benutzer- und systemdefiniert fehlende Werte ausschließen

Behandelt die benutzerdefiniert fehlenden Werte als gültige Werte. Dies ist die Standardeinstellung.

##### Benutzerdefiniert fehlende Werte sind gültig

Ignoriert die benutzerdefiniert fehlenden Wertbezeichnungen und behandelt sie als gültige Werte.

#### Intervallbehandlung

Eine Option, um zu steuern, wie mit Datensätzen umgegangen werden soll, deren Intervall mit der Anfangs- und Endzeit in Konflikt steht. Sie wird wirksam, wenn zwei Zeitvariablen mit einer Intervallvariablen im Hauptdialog angegeben sind.

##### Datensätze mit Konflikten verwerfen

Löscht die gesamten seriellen Datensätze des Subjekts, wenn der Intervallwert mit der Anfangs- und Endzeit in Konflikt steht. Dies ist die Standardeinstellung.

##### Intervallwerte auf der Basis der Start- und Endzeit ermitteln

Erkennt den Intervallwert aus der Anfangs- und Endzeit.

## Parametrische gemeinsam genutzte Frailty-Modelle: Modell

### Modell

Eine optionale Anzeige zum Angeben der Modelloptionen und -einstellungen.

#### Verteilung der Überlebenszeit

Eine Option zur Angabe der Verteilung der Überlebenszeit.

##### Weibull

Gibt die Weibull-Verteilung an. Dies ist die Standardeinstellung.

##### Exponentiell

Gibt die Exponentialverteilung an.

##### Logarithmische Normalverteilung

Gibt die Log-Normal-Verteilung an.

##### Log-logistisch

Gibt die Log-logistische Verteilung an.

#### Kovariateneinstellungen

Kovariatenvariablen angeben.

#### Faktoreinstellungen

Geben Sie Faktorvariablen an.

### **Anfangswert des konstanten Terms**

Eine Option zur Angabe des Anfangswerts des konstanten Terms. Wenn er angegeben wird, muss er ein einzelner numerischer Wert sein und darf nicht 0 sein.

### **Anfangswert des Skalenparameters**

Eine Option zur Steuerung der Einstellung des Skalenparameters.

### **Standardfehler der entsprechenden OLS-Regression**

Verwendet den Standardfehler der entsprechenden gewöhnlichen Regression der kleinsten Quadrate als Anfangswert.

### **Invertiert den Standardfehler der entsprechenden OLS-Regression**

Verwendet den reziproken Wert des Standardfehlers.

### **Vom Benutzer angegebener Wert**

Wenn ein einzelner numerischer Wert angegeben wird, wird der Wert als Anfangswert verwendet. Falls angegeben, muss er größer als 0 sein.

### **Frailty-Komponente**

Ein optionaler Parameter zur Angabe der **Verteilung** der Fragmentierungskomponente.

#### **Gamma**

Gibt die Gammaverteilung an. Dies ist die Standardeinstellung.

#### **Inverse Normalverteilung**

Gibt die inverse Normalverteilung an.

#### **Anfangswert der Varianz**

Gibt den Anfangswert der Varianz der Fragmentierungskomponente an. Es muss ein einzelner numerischer Wert größer als 0 sein. Der Standardwert ist 1.0 für die Gammaverteilung und 0.1 für die inverse Normalverteilung.

## **Parametrische gemeinsam genutzte Frailty-Modelle: Schätzung**

### **Schätzung**

Eine optionale Anzeige zum Angeben der Einstellungen zum Steuern der Schätzung der gemeinsam genutzten Frailty-Modelle und des Auswahlprozesses für optionale Features.

### **Methode mit alternierender Richtung oder Multiplikatoren (ADMM)**

#### **Schnell**

Wendet die schnelle Wechselrichtungsmethode von Multiplikatoren (ADMM) an. Dies ist die Standardeinstellung.

#### **Traditionell**

Wendet den traditionellen ADMM-Algorithmus an.

### **L-1-Regularisierung anwenden**

Führt den Prozess zur Steuerung der Merkmalauswahl durch Das Feld **Penalty Parameter** gibt den Penalisierungsparemeter an, der den Regularisierungsprozess steuert. Der Wert muss ein Einzelwert größer als 0 sein. Die Standardeinstellung ist 0.001.

### **Modellkonvergenzkriterien**

#### **Parameterkonvergenz**

Gibt die Konvergenzkriterien für den Parameter an Es muss ein einzelner numerischer Wert sein, der zu  $[0, 1)$  gehört. Die Standardeinstellung ist 0,000001. Als **Typ** können Sie entweder **ABSOLUTE** auswählen, um die absolute Konvergenz auf die innere Optimierung anzuwenden, oder **RELATIVE**, um die relative Konvergenz auf die innere Optimierung anzuwenden. Der optionale **Wert** gibt einen numerischen Schwellenwert für den Konvergenztyp an.

#### **Konvergenz der Zielfunktion**

Gibt die Konvergenzkriterien für die Zielfunktion an Es muss ein einzelner numerischer Wert sein, der zu  $[0, 1)$  gehört. Die Standardeinstellung ist 0, d. h., die Konvergenzkriterien werden nicht angewendet. Als **Typ** können Sie entweder **ABSOLUTE** auswählen, um die absolute Konvergenz

auf die innere Optimierung anzuwenden, oder **RELATIVE**, um die relative Konvergenz auf die innere Optimierung anzuwenden. Der optionale **Wert** gibt einen numerischen Schwellenwert für den Konvergenztyp an.

#### **Konvergenz der Hesse-Matrix**

Gibt die Konvergenzkriterien für die Hessische Matrix an. Es muss ein einzelner numerischer Wert sein, der zu  $[0, 1)$  gehört. Die Standardeinstellung ist 0, d. h., die Konvergenzkriterien werden nicht angewendet. Als **Typ** können Sie entweder **ABSOLUTE** auswählen, um die absolute Konvergenz auf die innere Optimierung anzuwenden, oder **RELATIVE**, um die relative Konvergenz auf die innere Optimierung anzuwenden. Der optionale **Wert** gibt einen numerischen Schwellenwert für den Konvergenztyp an.

#### **Konvergenzkriterien für Residuen**

Eine Option zur Steuerung des Optimierungsprozesses.

##### **Sowohl primales als auch duales Residuum**

Wendet sowohl das ursprüngliche als auch das doppelte Konvergenzkriterium für Residuen an. Diese Einstellung ist standardmäßig.

##### **Nur primales Residuum**

Wendet das Konvergenzkriterium der ursprünglichen Residuen an.

##### **Nur duales Residuum**

Wendet das doppelte Konvergenzkriterium für Residuen an

#### **Methode**

Ein optionaler Parameter zur Angabe der Schätzmethode.

##### **Automatisch**

Wählt die Methode automatisch basierend auf dem Stichprobendataset aus. Diese Methode ist standardmäßig ausgewählt. Das Feld **Schwellenwertanzahl der Prädiktoren** gibt den Schwellenwert für die Anzahl der Prädiktoren an und muss eine einzelne ganze Zahl größer als 1 sein. Der Standardwert beträgt "1000".

##### **Newton-Raphson**

Wendet die Newton-Raphson-Methode an.

##### **L-BFGS**

Wendet den BFGS-Algorithmus mit begrenztem Speicher an. Das Feld **Aktualisieren** gibt die Anzahl vergangener Aktualisierungen an, die vom BFGS-Algorithmus mit begrenztem Speicher verwaltet werden, und muss eine einzelne ganze Zahl größer-gleich 1 sein. Der Standardwert ist 5.

#### **Iteration**

##### **Maximalzahl der Iterationen**

Gibt die maximale Anzahl der Iterationen an. Es muss eine einzelne ganze Zahl sein, die zu  $[1, 300]$  gehört. Die Standardeinstellung ist 20.

##### **Maximale Schritthalbierung**

Gibt die maximale Anzahl der Schritthalbierung an. Es muss eine einzelne ganze Zahl sein, die zu  $[1, 200]$  gehört. Die Standardeinstellung ist 5.

##### **Maximale Anzahl von Zeilensuchen**

Gibt die maximale Anzahl von Zeilensuchen an. Es muss eine einzelne ganze Zahl sein, die zu  $[1, 300]$  gehört. Die Standardeinstellung ist 20.

##### **Absolute Konvergenz für Iterationsprozess**

Gibt die absolute Konvergenz für den äußeren Iterationsprozess an. Es muss sich um einen einzelnen numerischen Wert handeln, der zu  $(0, 1)$  gehört. Die Standardeinstellung ist 0.0001.

##### **Relative Konvergenz für Iterationsprozess**

Gibt die relative Konvergenz für den äußeren Iterationsprozess an. Es muss sich um einen einzelnen numerischen Wert handeln, der zu  $(0, 1)$  gehört. Die Standardeinstellung ist 0.01.

## Parametrische gemeinsam genutzte Frailty-Modelle: Drucken

### Drucken

Eine optionale Anzeige, die die Tabellenausgaben steuert.

#### Details zur Faktorcodierung

Bei Auswahl dieser Option werden die Codierungsdetails der Faktoren angezeigt und gedruckt. Der Prozess wird ignoriert, wenn keine Faktoren wirksam sind.

#### Anfangswerte, die den Regressionsparametern zugeordnet sind

Wenn diese Option ausgewählt ist, werden die Anfangswerte angezeigt, die im Schätzprozess verwendet werden.

#### Verlauf der Modelliteration

Bei Auswahl dieser Option wird der Iterationsverlauf der Überlebensanalyse angezeigt. Geben Sie im Feld **Anzahl Schritte** die Anzahl der Schritte zwischen 1 und 99999999 an. Die Standardeinstellung ist 1.

## Parametrische gemeinsam genutzte Frailty-Modelle: Vorhersagen

### Vorhersage

Eine optionale Anzeige zum Bewerten und Speichern der vorhergesagten Statistiken im aktiven Dataset.

#### Zeitwerte für Scoring

##### Durch abhängige Variable(n) definierte Zeitwerte

Bewertet die **Vorhersagen** basierend auf der Zeitvariablen, die für das parametrische Überlebensmodell angegeben wurde

##### Regelmäßige Intervalle

Bewertet die **Vorhersagen** basierend auf zukünftigen Zeitwerten. Das Feld **Zeitintervall** gibt das Zeitintervall an und muss ein einzelner numerischer Wert größer als 0 sein. Das Feld **Anzahl Zeiträume** gibt die Anzahl der Zeiträume an und muss eine einzelne numerische Ganzzahl zwischen 2 und 100 sein.

##### Zeitdifferenz

Bewertet die **Vorhersagen** basierend auf der Zeitdauer, um die zukünftigen Zeitwerte zu definieren. Es muss sich um eine einzelne numerische Variable handeln.

#### Vorhersagen

##### Überleben

Bewertet und speichert die vorhergesagten Überlebensstatistiken im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredSurvival`.

##### Gefahr

Bewertet und speichert die vorhergesagten Gefahren im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredHazard`.

##### Kumulierte Gefahr

Bewertet und speichert die vorhergesagten kumulativen Gefahren im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredCumHazard`.

##### Bedingtes Überleben

Bewertet und speichert die vorhergesagten bedingten Überlebensstatistiken im aktiven Dataset. Der Standardname der benutzerdefinierten Variablen (oder Stammname) lautet `PredConditionalSurvival`. Der Prozess wird ignoriert, wenn `PASTTIME` nicht angegeben ist. Ein Wert für **Vergangene Überlebenszeit** ist erforderlich und gibt die vergangenen Zeitwerte für das Scoring an. Es muss sich um eine einzelne numerische Variable handeln.

##### Unbedingtes Überleben

Bewertet und speichert die vorhergesagten unbedingten Überlebensstatistiken im aktiven Dataset. Das Schlüsselwort wird standardmäßig unterdrückt. Falls angegeben, könnte darauf ein opti-



onaler, vom Benutzer angegebener Variablenname (oder Stammmname) in Klammern folgen. Der Standardname ist `PredUnCondSurvival`.

### **Unbedingtes Risiko**

Bewertet und speichert die vorhergesagten unbedingten Hazardstatistiken im aktiven Dataset. Das Schlüsselwort wird standardmäßig unterdrückt. Falls angegeben, könnte darauf ein optionaler, vom Benutzer angegebener Variablenname (oder Stammmname) in Klammern folgen. Der Standardname ist `PredUnCondHazard`.

### **Unbedingte cum-Gefahr**

Bewertet und speichert die vorhergesagten unbedingten kumulativen Hazardstatistiken im aktiven Dataset. Das Schlüsselwort wird standardmäßig unterdrückt. Falls angegeben, kann auf sie ein optionaler, vom Benutzer angegebener Variablenname (oder ein in Klammern angegebener Stammmname) folgen. Der Standardname ist `PredUnCondCumHazard`.

## **Parametrische gemeinsam genutzte Frailty-Modelle: Diagramm**

### **Diagramm**

#### **Funktionsdiagramme**

Eine Option zur Steuerung der Funktionsdiagramme.

#### **Typ**

##### **Überleben**

Erstellt das Diagramm für die unbedingten Überlebensfunktionen.

##### **Gefahr**

Erstellt das Diagramm für die unbedingten Hazardfunktionen.

##### **Dichte**

Erstellt ein Diagramm für die Dichtefunktionen.

#### **Anzahl anzuzeigender Punkte**

Gibt die Anzahl der Funktionspunkte zwischen 1 und 200 an. Die Standardeinstellung ist 100.

#### **Kovariatenwerte für Diagramm**

Eine Option zum Angeben der benutzerdefinierten Werte und zum Zuweisen dieser Werte zu den Prädiktoren. Standardmäßig werden die designierten Diagramme am Mittelwert jeder wirksamen Kovariate erstellt. Bei Angabe werden die designierten Diagramme basierend auf der Einstellung des Musters erstellt. Wenn doppelte Variablen vorhanden sind, wird die zuerst angegebene Variable erkannt und der Rest ignoriert. Ein Modelleffekt muss eine gültige Variable enthalten. Bei einer Kovariate muss der benutzerdefinierte Wert numerisch sein. Die Auslassung einer Variablen in Kraft gibt an, dass der Mittelwert standardmäßig für die Kovariate verwendet wird. Wenn einer Variablen ein ungültiger Wert zugeordnet wird, wird das angeforderte Muster nicht dargestellt.

#### **Faktorwerte für Diagramm**

Eine Option zum Angeben der benutzerdefinierten Werte und zum Zuweisen dieser Werte zu den Prädiktoren. Standardmäßig werden die designierten Diagramme mit der Kategorienhäufigkeit jedes Faktors in Kraft erstellt. Bei Angabe werden die designierten Diagramme basierend auf der Einstellung des Musters erstellt. Wenn doppelte Variablen vorhanden sind, wird die zuerst angegebene Variable erkannt und der Rest ignoriert. Ein Modelleffekt muss eine gültige Variable enthalten. Die fehlende Angabe einer Variablen gibt an, dass die Kategorienfrequenz standardmäßig für den Faktor verwendet wird. Wenn einer Variablen ein ungültiger Wert zugeordnet wird, wird das angeforderte Muster nicht dargestellt.

#### **Separate Zeilen für**

Eine Option zur Angabe einer kategorialen Variablen, mit der die Liniendiagramme gezeichnet werden.

#### **Maximale Anzahl von Linien in einem Diagramm**

Gibt die maximale Anzahl der Linien in einem Diagramm an, wenn **Separate Linien für** angegeben wird. Die Standardeinstellung ist 10.

# Parametrische gemeinsam genutzte Frailty-Modelle: Exportieren

## Exportieren

Wählen Sie **Modellinformationen in XML-Datei exportieren** aus, um die Modell- und Parameterinformationen zum Scoring in eine PMML-Datei zu schreiben. Sie müssen das Verzeichnis und den Dateinamen der PMML-Datei angeben, die gespeichert werden soll.

# Parametrische gemeinsam genutzte Frailty-Modelle: Ereignisse definieren

Eine Option zum Definieren des Status. Wenn die Statusvariable weggelassen wird, wird der Fehler oder das Ereignis zum Standardstatus für alle Fälle.

1. Wählen Sie im Menü die Option

**Analyse > Überleben > Parametrische gemeinsame Frailty-Modelle ...**

- 2.

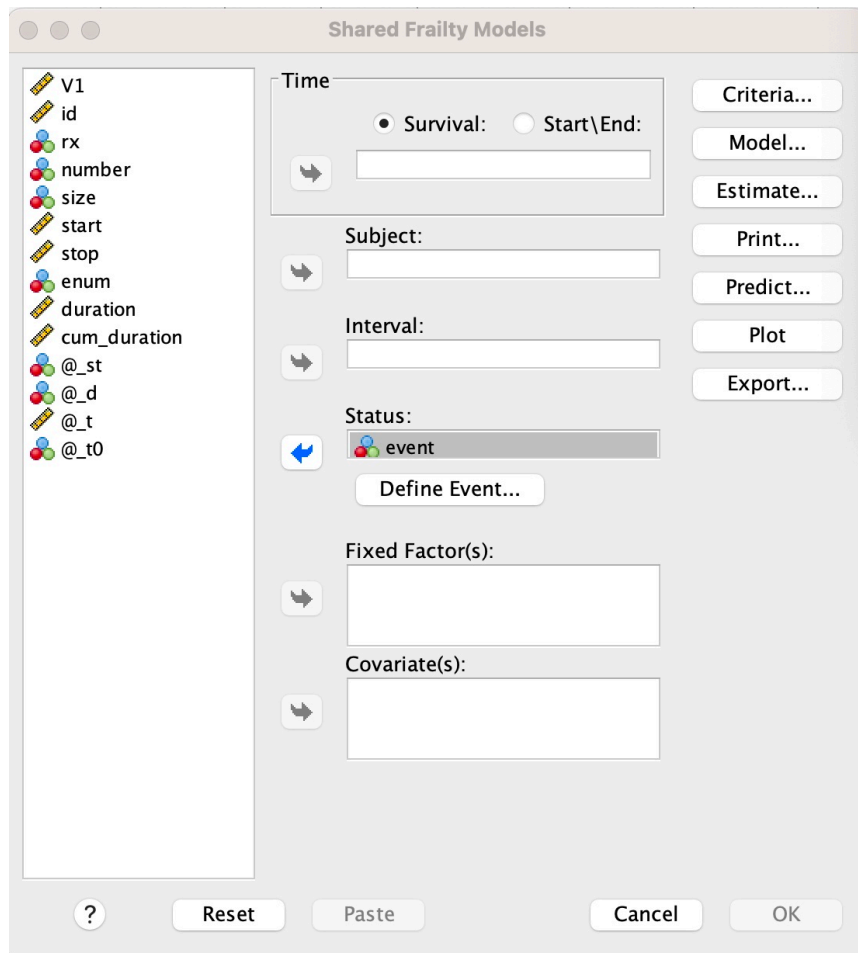


Abbildung 3. Gemeinsam genutzte Frailty-Modelle-Dialogfenster-Status

- 3.

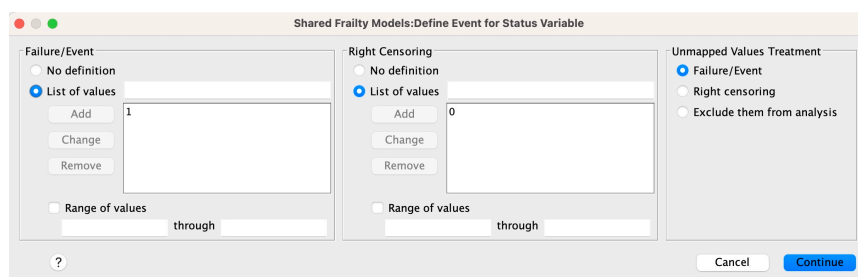


Abbildung 4. Gemeinsam genutzte Fragmentierungsmodelle-Dialogfenster-Status-Ereignis definieren

## Parametrische gemeinsam genutzte Frailty-Modelle-Beispiele

### Beispiel 1

```
SURVREG RECURRENT Y MIT x1 BY x2  
/MODEL SUBJECT = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL.
```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von y an eine Kovariate x1 und Faktor x2.

Die Überlebenszeit wird durch eine einzelne Variable y dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

Es wird angenommen, dass die Überlebenszeit einer Weibull-Verteilung folgt.

Es wird angenommen, dass die Varianz der Frailty einer Gammaverteilung folgt.

Alle gültigen Datensätze werden in der Überlebensanalyse verwendet.

### Beispiel 2

```
SURVREG RECURRENT Y MIT x1 BY x2  
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von y an eine Kovariate x1 und Faktor x2.

Die Überlebenszeit wird durch zwei Variablen y1 und y2 dargestellt, die die Start- und Endzeit angeben.

Die Subjekte werden durch die Variablen-ID identifiziert.

Es wird angenommen, dass die Überlebenszeit einer Log-Normalverteilung folgt.

Es wird angenommen, dass die Varianz der Frailty einer inversen Normalverteilung folgt.

Zeitintervalle werden durch die Variable z definiert. Für jedes Subjekt verwendet die Prozedur nur die Datensätze ohne Konflikte und schließt alle Datensätze nach dem ersten Fehlerstatus aus der Analyse aus.

### Beispiel 3

```
SURVREG RECURRENT y1 y2 WITH x1 BY x2(1)  
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC  
/STATUS VARIABLE=Ereignis FAILURE=1 RIGHT=0.
```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von y an eine Kovariate x1 und Faktor x2. Überleben

Zeit wird mit zwei Variablen y1 und y2 dargestellt, die die Start- bzw. Endzeit angeben. Für den Faktor x2 wird die Warengruppe "1" als zu modellierende Baseline festgelegt.

Die Subjekte werden durch die Variablen-ID identifiziert.

Es wird angenommen, dass die Überlebenszeit einer Log-logistischen Verteilung folgt.

Es wird angenommen, dass die Varianz der Frailty einer inversen Normalverteilung folgt.

Das Variablenereignis wird angegeben, um den Status mit 1 und 0 für Fehler bzw. Rechtszensierung zu definieren.

### Beispiel 4

```

SURVREG RECURRENT Y MIT x1 BY x2
/MODEL SUBJECT = ID
/STATUS VARIABLE=Ereignis FAILURE=1 RIGHT=0
/PREDICT UNCONDSURVIVAL UNCONDHAZARD UNCONDCUMHAZARD
/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).

```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von  $y$  an eine Kovariate  $x_1$  und Faktor  $x_2$ . Die Überlebenszeit wird durch eine einzelne Variable  $y$  dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

Unbedingtes oder befüllungsbasiertes Überleben, Hazard und kumulative Hazard-Rate werden bewertet und im aktiven Dataset gespeichert.

Die Kurven für unbedingtes oder auf der Bevölkerung basierendes Überleben und Hazard werden durch die Kategorien in  $x_2$  getrennt dargestellt.

### Beispiel 5

```

SURVREG RECURRENT Y MIT x1 BY x2
/MODEL SUBJECT = id FRAILITY=GAMMA DISTRIBUTION=WEIBULL
/STATUS VARIABLE=Ereignis FAILURE=1 RIGHT=0
/SCHÄTZUNG HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0 SELECTFEATURES=TRUE
PENALTY=0.01.

```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von  $y$  an eine Kovariate  $x_1$  und Faktor  $x_2$ . Die Überlebenszeit wird durch eine einzelne Variable  $y$  dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

Es wird angenommen, dass die Überlebenszeit einer Weibull-Verteilung folgt.

Es wird angenommen, dass die Varianz der Frailty einer Gammaverteilung folgt.

Die Konvergenzkriterien basieren auf der Hessischen Matrix. Sie verwendet  $1e-12$  als relative Konvergenz.

Das Modell enthält einen Penalisierungsterm zur Steuerung des Regularisierungsprozesses. Der Penalisierungsparameter ist auf  $0.01$  gesetzt.

### Beispiel 6

```

SURVREG RECURRENT Y MIT x1 BY x2
/MODEL SUBJECT = ID
/STATUS VARIABLE=infect FAILURE=1 RIGHT=0
/SCHÄTZUNG MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.

```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von  $y$  an eine Kovariate  $x_1$  und Faktor  $x_2$ . Die Überlebenszeit wird durch eine einzelne Variable  $y$  dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

Die Prozedur gibt die maximale Anzahl der Zeilensuche auf 100, die maximale Anzahl der Iterationen auf 50 und die maximale Anzahl der Schritthalbierung auf 20 an.

# Parametrische gemeinsam genutzte Frailty-Modelle-Fallstudie für Rekurrent-Daten

Parametrische gemeinsam genutzte Frailty-Modelle-Fallstudie für Rekurrent-Daten

Anwendungsfallname-Nebeneffekt der Behandlung.

Akteure-Ermittler und Praktiker der öffentlichen Gesundheit.

Vorbedingungen-Ein bereinigtes Dataset, das auf der Überlebenszeit, dem Nebeneffektstatus und den zu anpassenden Prädiktoren basiert.

Beschreibung-Patrick, ein Ermittler des öffentlichen Gesundheitswesens, untersucht eine Datenstichprobe mit 20 Teilnehmern. Diese Teilnehmer werden in einer Studie über eine leichte Nebenwirkung rekrutiert, die potenziell durch eine neue Behandlung verursacht wird. Der Behandlungsdesigner behauptet, dass es keine Unterschiede zwischen Männern und Frauen in Bezug auf die Nebenwirkung geben würde. Patrick möchte eine solche Hypothese bewerten. Die in der Datenstichprobe enthaltenen Variablen werden wie folgt aufgelistet:

- patID: ID-Nummer zur Identifizierung eines eindeutigen Teilnehmers.
- endTime: Überlebenszeit (in Tagen) der Nebenwirkung nach einer Behandlung, die vom Beginn einer Behandlung bis zu einem berichteten Nebeneffekt oder einer Zensur innerhalb von 60 Tagen gemessen wird.
- sideEffect: Nebeneffektstatus, Status = 0, wenn zensiert, und Status = 1, wenn der milde Seiteneffekt gemeldet wird
- Alter: Das Alter des Teilnehmers im Forschungszeitraum.
- weiblich: weiblich = 0, wenn männlich und weiblich = 1, wenn weiblich.

Es können mehrere Verfahren angewendet werden, was dazu führt, dass mehrere Datensätze von Wiederholungszeiten für einen bestimmten Teilnehmer gemessen werden. Die Startzeit ist immer 0 für jeden Datensatz, der in der Datenstichprobe weggelassen wird. Patrick ist daran interessiert, die Überlebens- und Gefahrenfunktionen zu visualisieren, um einen Vergleich zwischen einem Mann und einer Frau zu ziehen, indem er ihr Alter und ihre Gebrechlichkeit kontrolliert. Er ist sich bewusst, dass die Behandlungen, die demselben Teilnehmer verabreicht werden, eher korreliert sind. Unter der Annahme, dass die Überlebenszeit einer Weibull-Verteilung folgt, beschließt Patrick, in SPSS Statistics ein parametrisches Überlebensmodell für gemeinsam genutzte Frailty zu erstellen, um die Behandlungsabhängigkeit für denselben Teilnehmer zu berücksichtigen.

Syntax-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
```

```

15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILITY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .

```

## Übersicht:

Die von Patrick angegebene Syntax bezeichnet endTime als einzelne abhängige Zeitvariable. Die Prozedur geht automatisch davon aus, dass die Startzeit für jeden Datensatz 0 ist. Die Variablen "age" und "female" werden als Kovariate bzw. als Faktor modelliert. Es wird angenommen, dass die Rezidivüberlebenszeiten einer Weibull-Verteilung folgen. Es wird angenommen, dass der nicht beobachtete Frailty-Term einer Gammaverteilung folgt, und seine Varianzkomponente wird modelliert. In Bezug auf die Ausgaben enthält die Tabelle "Modellzusammenfassung" die Prozedur- und Modellinformationen. Die Übersichtstabelle 'Fallverarbeitung' enthält eine umfassende Liste der Fehler-/Zensurstatus sowie der Fälle, die von der Analyse ausgeschlossen sind.

In Patricks Datenstichprobe sind alle Datensätze gültig und in die Analyse eingeschlossen. Durch den Vergleich der Log-Likelihood mit dem des entsprechenden Modells ohne die Frailty-Komponente erreicht das Modell für gemeinsam genutzte Frailty keine signifikante Stufe ( $p$ -Wert = 0.168). Patrick fragt sich, ob es notwendig ist, einen Shared-Frailty-Begriff in das Modell aufzunehmen. Der geschätzte Beschleunigungsfaktor eines männlichen Teilnehmers ist 1.017, der durch Berechnung des Exponenten des geschätzten Regressionskoeffizienten 0.017 von [weiblich = 0.0] berechnet wird. Das zugehörige 95 %-Konfidenzintervall ist (0,688, 1.504). Diese Ergebnisse deuten darauf hin, dass eine männliche Person hat fast den gleichen Beschleunigungsfaktor wie eine weibliche Person mit dem gleichen Alter und Gebrechlichkeit. Auf Bevölkerungsebene stellt Patrick die unbedingten Überlebens- und Hazardkurven getrennt für Männer und Frauen dar, die am Stichprobenmittelwert des Alters (37.45 Jahre alt) bewertet werden.

Patrick bestätigt, dass für jeden festen Wert der Überlebenszeit ein Männchen und ein Weibchen im Durchschnitt dieselbe Überlebenswahrscheinlichkeit haben sollen. Interessanterweise stellt Patrick trotz einer unimodalen Form, die in dem unbedingten Gefahrendiagramm gezeigt wird, fest, dass innerhalb von 60 Tagen die Populationsgefahr tatsächlich zunimmt. Dieses Verhalten kann das Vorhandensein des Fragmentierungseffekts implizieren. Um die durch die Behandlungen verursachte Nebenwirkung weiter zu untersuchen, kann Patrick mit einem Modell ohne die frailty Komponente fortfahren und das Verhalten von Männern und Frauen vergleichen. Darüber hinaus kann er erwägen, mit den Teilnehmern für einen Zeitraum von mehr als 60 Tagen zu folgen, um mehr Daten zu sammeln.

## Bemerkungen

---

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden. IBM stellt dieses Material möglicherweise auch in anderen Sprachen zur Verfügung. Für den Zugriff auf das Material in einer anderen Sprache ist eine Kopie des Produkts oder der Produktversion in der jeweiligen Sprache erforderlich.

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

*IBM Director of Licensing*

*IBM Europe, Middle East & Africa*

*North Castle Drive, MD-NC119*

*Armonk, NY 10504-1785*

*France*

Bei Lizenzanforderungen zu Double-Byte-Information (DBCS) wenden Sie sich bitte an die IBM Abteilung für geistiges Eigentum in Ihrem Land oder senden Sie Anfragen schriftlich an folgende Adresse:

*Intellectual Property Licensing*

*Legal and Intellectual Property Law*

*IBM Japan Ltd.*

*19-21, Nihonbashi-Hakozakicho, Chuo-ku*

*Tokyo 103-8510, Japan*

Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you. -->

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter als IBM werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen

Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

*IBM Director of Licensing*

*IBM Europe, Middle East & Africa*

*North Castle Drive, MD-NC119*

*Armonk, NY 10504-1785*

*France*

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des im Dokument aufgeführten Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Die angeführten Leistungsdaten und Kundenbeispiele dienen nur zur Illustration. Die tatsächlichen Ergebnisse beim Leistungsverhalten sind abhängig von der jeweiligen Konfiguration und den Betriebsbedingungen.

Alle Informationen zu Produkten anderer Anbieter als IBM stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussage zu Leistung, Kompatibilität und anderen Merkmalen von Produkten anderer Anbieter als IBM machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter als IBM sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufes. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

#### COPYRIGHTLIZENZ:

Diese Veröffentlichung enthält Beispielanwendungsprogramme, die in Quellsprache geschrieben sind und Programmieretechniken in verschiedenen Betriebsumgebungen veranschaulichen. Sie dürfen diese Beispielprogramme ohne Zahlung an IBM kopieren, ändern und verteilen, wenn dies zu dem Zweck geschieht, Anwendungsprogramme zu entwickeln, zu verwenden, zu vermarkten oder zu verteilen, die mit der Anwendungsprogrammierschnittstelle für die Betriebsumgebung konform sind, für die diese Beispielprogramme geschrieben werden. Diese Beispiele wurden nicht unter allen denkbaren Bedingungen getestet. Daher kann IBM die Zuverlässigkeit, Wartungsfreundlichkeit oder Funktion dieser Programme weder zusagen noch gewährleisten. Die Beispielprogramme werden ohne Wartung (auf "as-is"-Basis) und ohne jegliche Gewährleistung zur Verfügung gestellt. IBM übernimmt keine Haftung für Schäden, die durch die Verwendung der Beispielprogramme entstehen.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

© Copyright IBM Corp. 2021. Teile des vorliegenden Codes wurden aus Beispielprogrammen der IBM Corp. abgeleitet.

© Copyright IBM Corp. 1989 - 2021. All rights reserved.

## Marken

---

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corp in den USA und/ oder anderen Ländern. Weitere Produkt- oder Servicenamen können Marken von IBM oder anderen Her-



stellern sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite "Copyright and trademark information" unter [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, das Adobe-Logo, PostScript und das PostScript-Logo sind Marken oder eingetragene Marken der Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder ihrer Tochtergesellschaften in den USA oder anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind eingetragene Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.



---

# Index

## A

- Abweichungsfunktion
  - zum Schätzen des Skalierungswerts für Streuung [9](#)
- Anpassungsgüte
  - in multinomialer logistischer Regression [8](#)
- Asymptotische Regression
  - in nicht linearer Regression [21](#)

## B

- Benutzerdefinierte Modelle
  - in multinomialer logistischer Regression [7](#)
- Binär logistische Regression [1](#)
- Binäre logistische Regression [2](#)

## C

- Chi-Quadrat nach Pearson
  - Anpassungsgüte [8](#)
- Cook-D
  - in logistischer Regression [4](#)

## D

- Delta
  - als Korrektur für Zellen mit null Beobachtungen [9](#)
- DfBeta
  - in logistischer Regression [4](#)
- Dichtemodell
  - in nicht linearer Regression [21](#)

## E

- Ertragsdichtemodell
  - in nicht linearer Regression [21](#)

## F

- Fiduzialkonfidenzintervalle
  - in Probit-Analyse [12](#)

## G

- Gauß-Modell
  - in nicht linearer Regression [21](#)
- Gesättigte Modelle
  - in multinomialer logistischer Regression [7](#)
- Gewichtungsschätzung
  - ANOVA und Schätzungen anzeigen [25](#)
  - Beispiel [23](#)
  - beste Gewichtung als neue Variable speichern [25](#)
  - Iterationsverlauf [25](#)
  - Log-Likelihood [23](#)
  - Statistik [23](#)
  - zusätzliche Funktionen beim Befehl [25](#)

- Gompertz-Modell
  - in nicht linearer Regression [21](#)

## H

- Hebelwerte
  - in logistischer Regression [4](#)
- Hosmer-Lemeshow-Anpassungsstatistik
  - in logistischer Regression [5](#)

## I

- Iteration
  - in logistischer Regression [5](#)
  - in Probit-Analyse [12](#)
- Iterationen
  - in multinomialer logistischer Regression [9](#)
- Iterationsverlauf
  - in multinomialer logistischer Regression [9](#)

## J

- Johnson-Schumacher-Modell
  - in nicht linearer Regression [21](#)

## K

- kategoriale Kovariaten [4](#)
- Kernel Ridge
  - alpha [30](#)
  - coef0 [30](#)
  - Einzelmodell [30](#)
  - Gamma [30](#)
  - Grad [30](#)
  - Modellauswahl [30](#)
- Kernel-Ridge-Regression
  - Anzeige [32](#)
  - Grafik [32](#)
  - Kreuzvalidierungsfalten [32](#)
  - Parameter [32](#)
  - Rasterparameter [32](#)
  - speichern [32](#)
- Klassifikation
  - in multinomialer logistischer Regression [6](#)
- Klassifikationstabellen
  - in multinomialer logistischer Regression [8](#)
- Konfidenzintervalle
  - in multinomialer logistischer Regression [8](#)
- konstanter Term
  - in "Lineare Regression" [5](#)
- Konstanter Term
  - einschließen oder ausschließen [7](#)
- Kontraste
  - in logistischer Regression [4](#)
- Konvergenzkriterium

Konvergenzkriterium (*Forts.*)  
in multinomialer logistischer Regression [9](#)  
Korrelationsmatrix  
in multinomialer logistischer Regression [8](#)  
Kovarianzmatrix  
in multinomialer logistischer Regression [8](#)  
Kovariaten  
in logistischer Regression [4](#)

## L

Likelihood-Quotient  
Anpassungsgüte [8](#)  
zum Schätzen des Skalierungswerts für Streuung [9](#)  
Lineare Regression  
Gewichtungsschätzung [23](#)  
Zweistufige Regressionsmethode der kleinsten Quadrate [25](#)  
Log-Likelihood  
in Gewichtungsschätzung [23](#)  
in multinomialer logistischer Regression [8](#)  
Log-modifiziertes Modell  
in nicht linearer Regression [21](#)  
Logistische Regression  
Anzeigeoptionen [5](#)  
Auswahlmethoden für Variablen [3](#)  
Auswahlregel definieren [3](#)  
Beispiel [2](#)  
binär [1](#)  
Einflussmaße [4](#)  
Hosmer-Lemeshow-Anpassungsstatistik [5](#)  
Iteration [5](#)  
kategoriale Kovariaten [4](#)  
Klassifikationstrennwert [5](#)  
Koeffizienten [2](#)  
konstanter Term [5](#)  
Kontraste [4](#)  
Regel definieren [3](#)  
Residuen [4](#)  
Speichern von neuen Variablen [4](#)  
Statistik [2](#)  
Statistiken und Diagramme [5](#)  
Vorhergesagte Werte [4](#)  
Wahrscheinlichkeit für schrittweise Methode [5](#)  
Zeichenfolge-Kovariaten [4](#)  
zusätzliche Funktionen beim Befehl [6](#)  
Logistische Regressionsanalyse [2](#)

## M

McFadden-R-Quadrat  
in multinomialer logistischer Regression [8](#)  
Metcherlich-Gesetz der abnehmenden Erträge  
in nicht linearer Regression [21](#)  
Michaelis-Menten-Modell  
in nicht linearer Regression [21](#)  
Mittlere effektive Dosis  
in Probit-Analyse  
[12](#)  
Modell mit Haupteffekten  
in multinomialer logistischer Regression [7](#)  
Morgan-Mercer-Florin-Modell  
in nicht linearer Regression [21](#)

Multinomiale logistische Regression  
Exportieren von Modellinformationen [10](#)  
Kriterien [9](#)  
Modelle [7](#)  
Referenzkategorie [8](#)  
speichern [10](#)  
Statistik [8](#)  
zusätzliche Funktionen beim Befehl [10](#)

## N

Nagelkerke-R-Quadrat  
in multinomialer logistischer Regression [8](#)  
Nebenbedingungen für Parameter  
in nicht linearer Regression [22](#)  
Nicht lineare Modelle  
in nicht linearer Regression [21](#)  
Nicht lineare Regression  
Ableitungen [22](#)  
Anfangswerte [20](#)  
Bedingungslogik [20](#)  
Beispiel [19](#)  
Bootstrap-Schätzungen [22](#)  
gängige nicht lineare Modelle [21](#)  
Interpretieren von Ergebnissen [23](#)  
Levenberg-Marquardt-Algorithmus [22](#)  
Nebenbedingungen für Parameter [22](#)  
neue Variablen speichern [22](#)  
Parameter [20](#)  
Residuen [22](#)  
Schätzmethoden [22](#)  
segmentiertes Modell [20](#)  
sequenzielle quadratische Optimierung [22](#)  
Statistik [19](#)  
Verlustfunktion [21](#)  
Vorhergesagte Werte [22](#)  
zusätzliche Funktionen beim Befehl [23](#)

## P

Parallelitätstest  
in Probit-Analyse  
[12](#)  
Parameterschätzungen  
in multinomialer logistischer Regression [8](#)  
Parametrische beschleunigte Fehlerzeitmodelle  
drucken [36](#)  
Export [38](#)  
Kriterien [34](#)  
modell [34](#)  
plotten [38](#)  
Schätzung [35](#)  
und Einrichtungen [33](#)  
vorhersagen [37](#)  
Parametrische Frailty-Modelle  
Statusvariablen für Überlebensverteilung [46](#)  
Parametrische gemeinsam genutzte Frailty-Modelle  
drucken [44](#)  
Export [46](#)  
Kriterien [41](#)  
modell [41](#)  
plotten [45](#)  
Schätzung [42](#)

Parametrische gemeinsam genutzte Frailty-Modelle (*Forts.*)  
und Einrichtungen [40](#)  
vorhersagen [44](#)

Peal-Reed-Modell

in nicht linearer Regression [21](#)

Pearson-Chi-Quadrat

zum Schätzen des Skalierungswerts für Streuung [9](#)

Probit-Analyse

Bereich definieren [12](#)

Fiduzialkonfidenzintervalle [12](#)

Iteration [12](#)

Kriterien [12](#)

Mittlere effektive Dosis [12](#)

natürliche Rücklaufquote [12](#)

Parallelitätstest [12](#)

Statistik [12](#)

zusätzliche Funktionen beim Befehl [12](#)

Probit-Regression

Beispiel [10](#)

Statistikdaten [10](#)

## Q

Quantilregression

-Modell [15](#)

Anzeige [16](#)

Beispiel: [12](#)

Export [18](#)

Kriterien [14](#)

Speichern [17](#)

Statistik [12](#)

## R

R-Quadrat nach Cox und Snell

in multinomialer logistischer Regression [8](#)

Referenzkategorie

in multinomialer logistischer Regression [8](#)

Regression mit Nebenbedingungen

in nicht linearer Regression [22](#)

Richards-Modell

in nicht linearer Regression [21](#)

Rückwärtselimination

in logistischer Regression [3](#)

## S

Schritthalbierungen

in multinomialer logistischer Regression [9](#)

Schrittweise Auswahl

in logistischer Regression [3](#)

in multinomialer logistischer Regression [7](#)

Singularität

in multinomialer logistischer Regression [9](#)

Skalierungswert für Streuung

in multinomialer logistischer Regression [9](#)

SPSS logistische Regression [2](#)

Sterbetafel

Statusvariablen für Überlebensverteilung [39](#)

## T

Tabellen mit Zellenwahrscheinlichkeiten

Tabellen mit Zellenwahrscheinlichkeiten (*Forts.*)  
in multinomialer logistischer Regression [8](#)

Trennung

in multinomialer logistischer Regression [9](#)

## U

Überleben AFT

Überlebensdialog-Kategorievariablen [39](#)

Überlebensanalyse

in der Kernel Ridge Regression [30](#)

## V

Verhältnis der 3. Potenzen

in nicht linearer Regression [21](#)

Verhältnis der Quadrate

in nicht linearer Regression [21](#)

Verhulst-Modell

in nicht linearer Regression [21](#)

Von-Bertalanffy-Modell

in nicht linearer Regression [21](#)

Vorwärtsselektion

in logistischer Regression [3](#)

## W

Weibull-Modell

in nicht linearer Regression [21](#)

## Z

Zeichenfolge-Kovariaten

in logistischer Regression [4](#)

Zellen mit null Beobachtungen

in multinomialer logistischer Regression [9](#)

Zweistufige Regressionsmethode der kleinsten Quadrate

Beispiel [25](#)

Instrumentalvariablen [25](#)

Kovarianz der Parameter [26](#)

Speichern von neuen Variablen [26](#)

Statistik [25](#)

zusätzliche Funktionen beim Befehl [26](#)





