

*IBM SPSS Regression 25*

**IBM**

**Hinweis**

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 27 gelesen werden.

**Produktinformation**

Diese Ausgabe bezieht sich auf Version 25, Release 0, Modifikation 0 von IBM® SPSS Statistics und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuauflage geändert wird.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs

*IBM SPSS Regression 25,*

herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2017

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:

TSC Germany

Kst. 2877

Juli 2017

---

# Inhaltsverzeichnis

<b>Regression</b> . . . . .	<b>1</b>
Auswählen einer Prozedur für binär logistische Re- gressionsmodelle . . . . .	1
Logistische Regression . . . . .	2
Logistische Regression: Regel definieren . . . . .	3
Logistische Regression: Methode zur Auswahl von Variablen . . . . .	3
Logistische Regression: Kategoriale Variablen defi- nieren . . . . .	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: Referenzkate- gorie . . . . .	8
Multinomiale logistische Regression: Statistiken.	8
Multinomiale logistische Regression: Konvergenz- kriterien . . . . .	9
Multinomiale logistische Regression: Optionen . . . . .	9
Multinomiale logistische Regression: Speichern	10
Zusätzliche Funktionen beim Befehl NOMREG	11
Probit-Analyse . . . . .	11
Probit-Analyse: Bereich definieren . . . . .	12
Probit-Analyse: Optionen . . . . .	12
Zusätzliche Funktionen beim Befehl PROBIT . . . . .	13
Nicht lineare Regression . . . . .	13
Bedingungslogik (nicht lineare Regression) . . . . .	14
Nicht lineare Regression: Parameter . . . . .	15
Nicht lineare Regression: Gängige Modelle . . . . .	15

Nicht lineare Regression: Verlustfunktion . . . . .	16
Nicht lineare Regression: Nebenbedingungen für Parameter . . . . .	16
Nicht lineare Regression: Neue Variablen spei- chern . . . . .	17
Nicht lineare Regression: Optionen . . . . .	17
Interpretieren von Ergebnissen der nicht linearen Regression. . . . .	17
Zusätzliche Funktionen beim Befehl NLR . . . . .	18
Gewichtungsschätzung . . . . .	18
Gewichtungsschätzung: Optionen . . . . .	19
Zusätzliche Funktionen beim Befehl WLS . . . . .	20
Zweistufige Regressionsmethode der kleinsten Qua- drate . . . . .	20
Zweistufige kleinste Quadrate: Optionen . . . . .	21
Zusätzliche Funktionen beim Befehl 2SLS . . . . .	21
Codierungsschemata für kategoriale Variablen. . . . .	21
Abweichung . . . . .	21
Einfach . . . . .	22
Helmert . . . . .	22
Differenz . . . . .	23
Polynomial . . . . .	23
Wiederholt . . . . .	23
Spezial . . . . .	24
Indikator . . . . .	25

<b>Bemerkungen.</b> . . . . .	<b>27</b>
Marken. . . . .	28
<b>Index</b> . . . . .	<b>31</b>



---

# 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, Einflussstatistiken 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** bietet die folgenden 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 Einflussstatistiken

**Multinomiale logistische Regression** bietet die folgenden exklusiven 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

**Anmerkung:** 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.

**Anmerkung:** 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 in den Menüs Folgendes aus:  
**Analysieren > Regression > Binär logistisch...**
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 klicken Sie dann auf **Regel**.

## Logistische Regression: Regel definieren

Die durch die Auswahlregel definierten Fälle werden in die Analyse 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 generiert. 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.

- *Einschluss*. 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)*. Rückwärtsgerichtete schrittweise Auswahl. Der Ausschlusstest basiert auf der Wahrscheinlichkeit der Likelihood-Quotienten-Statistik auf der Grundlage bedingter Parameterschätzungen.
- *Rückwärtselimination (Likelihood-Quotient)*. Rückwärtsgerichtete schrittweise Auswahl. Der Ausschlusstest basiert auf der Wahrscheinlichkeit der Likelihood-Quotienten-Statistik auf der Grundlage maximaler, partieller Likelihood-Schätzungen.
- *Rückwärtselimination (Wald)*. Rückwärtsgerichtete schrittweise Auswahl. 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 die Werte, die durch das Modell vorhergesagt werden. Die verfügbaren Optionen sind "Wahrscheinlichkeiten" und "Gruppenzugehörigkeit".

- *Wahrscheinlichkeiten.* Speichert für jeden Fall die vorhergesagte Wahrscheinlichkeit 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, dann wird die vorhergesagte Wahrscheinlichkeit von 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.* Das Analogon zur Einflussstatistik nach Cook in der logistischen Regression. 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.
- *Differenz in Beta.* Die Differenz im Beta-Wert entspricht der Änderung im 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 auf einer Logit-Skala vorhergesagten Fall. 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.* Der Quotient aus dem Residuum und einer Schätzung seiner Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- *Abweichung.* 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 für die Anpassungsgüte ist robuster als die herkömmliche Statistik für die Anpassungsgüte der logistischen Regression, insbesondere bei Modellen mit fortlaufenden Kovariaten und Untersuchungen mit kleinen Stichproben. 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 Aufnahmewert ist. Sie wird ausgeschlossen, 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, vorhergesagte 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 Informationen zur Syntax finden Sie in der 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 Kovarianzstruktur 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 in den Menüs Folgendes aus:  
**Analysieren > Regression > Multinomial logistisch...**
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 und 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 wurde. Dann wechselt der Algorithmus zwischen Rückwärtselimination bei den Ter-

men 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 wurde. 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.

## Erstellen von Termen

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.** Hier können Sie einen nicht negativen Wert kleiner als 1 angeben. Dieser Wert wird zu jeder leeren Zelle der Kreuztabelle von Antwortkategorie gegenüber Kovariatenstruktur addiert. 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 "Vorwärtseingabe" 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 Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

---

## Probit-Analyse

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-Analyse 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.

**Statistiken.** 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> (Gill, Murray, Saunders und Wright) zur Schätzung der Modellparameter vorgeschlagen und implementiert wurden.

Erläuterungen der Daten für die Probit-Analyse

**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 Regressi-

on. 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-Analyse

1. Wählen Sie in den Menüs Folgendes aus:

**Analysieren > Regression > Probit...**

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, klicken Sie im Dialogfeld auf **Bereich definieren**, um die 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-Listenfeld "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 effektiven Dosen 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.
- *Fiduzialkonfidenzintervalle*. 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.* Schätzt die natürliche Rücklaufquote aus den Stichprobendaten. Die Daten sollten einen Fall enthalten, der die Kontrollstufe darstellt, für die der Wert der Kovariate(n) 0 ist. Die natürliche Rücklaufquote wird durch Probit geschätzt, indem der Antwortanteil für die Kontrollstufe als Anfangswert verwendet wird.
- *Wert.* Legt die natürliche Rücklaufquote im Modell fest (wählen Sie diese Option aus, wenn Sie die natürliche Rücklaufquote im Voraus wissen). 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 Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

---

## 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 in der Form von  $Y = A + BX^2$  nicht notwendig ist. Durch Definieren von  $W = X^2$  erhalten Sie das einfache lineare Modell  $Y = A + BW$ , welches mit traditionellen Methoden wie der Prozedur "Lineare Regression" 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.

**Statistiken.** 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.

*Hinweis:* 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 in den Menüs Folgendes 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.

Nehmen Sie beispielsweise ein segmentiertes Modell an, das bei  $X \leq 0$  gleich 0, bei  $0 < X < 1$  gleich  $X$  und bei  $X \geq 1$  gleich 1 ist. Der Ausdruck hierfür lautet:

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

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  $0 * 0 + 1 * X + 0 * 1 = X$ .

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

Kompliziertere Beispiele können leicht gebildet werden, indem andere logische Ausdrücke und Ergebnisausdrücke eingesetzt werden. Denken Sie daran, dass doppelte Ungleichungen, beispielsweise  $0 < X < 1$ , als zusammengesetzte Ausdrücke aufgeführt werden müssen, in diesem Fall als  $(X > 0 \ \& \ X < 1)$ .

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	$b1 + b2 * \text{Exp}(b3 * x)$
Asymptotische Regression	$b1 - (b2 * (b3 ** x))$
Dichte	$(b1 + b2 * x) ** (-1 / b3)$
Gauß	$b1 * (1 - b3 * \text{Exp}(-b2 * x ** 2))$
Gompertz	$b1 * \text{Exp}(-b2 * \text{Exp}(-b3 * x))$
Johnson-Schumacher	$b1 * \text{Exp}(-b2 / (x + b3))$
Log-modifiziert	$(b1 + b3 * x) ** b2$
Log-logistisch	$b1 - \ln(1 + b2 * \text{Exp}(-b3 * x))$
Metcherlich-Gesetz der abnehmenden Erträge	$b1 + b2 * \text{Exp}(-b3 * x)$
Michaelis-Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1 + b2 * \text{Exp}(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$

Tabelle 1. Beispiele für Modellformeln: (Forts.)

Name	Modellformel
Verhältnis der 3. Potenzen	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
Verhältnis der Quadrate	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \text{Exp}(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \text{Exp}(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \text{Exp}(-b3 * x)) ** (1 / (1 - b4))$
Weibull (PP Plots Test Dist.)	$b1 - b2 * \text{Exp}(-b3 * x ** b4)$
Ertragsdichte	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

## 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. Zeichenfolgekonstanten 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  $\leq$ ,  $=$  oder  $\geq$ .
- 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.

- *Residuen*. Speichert Residuen unter dem Variablennamen "resid".
- *Vorhergesagte Werte*. Speichert vorhergesagte Werte unter dem Variablennamen "pred\_".
- *Ableitungen*. Es wird eine Ableitung pro Modellparameter gespeichert. Ableitungsnamen werden erstellt, indem das Präfix "d." den ersten sechs Zeichen der Parameternamen vorangestellt wird.
- *Werte der Verlustfunktion*. Diese Option steht nur beim Festlegen der eigenen Verlustfunktion zur Verfügung. 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 zum Schätzen des Standardfehlers einer Statistik, bei der wiederholte Stichproben aus dem Originaldataset verwendet werden. 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 steht für Modelle mit und ohne Nebenbedingungen zur Verfügung. 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*. Standardalgorithmus für Modelle ohne Nebenbedingungen. 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 Informationen zur Syntax 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 Custom Tables und Advanced Statistics). Wenn Ihre Daten nicht unabhängig sind, weil Sie beispielsweise dieselbe Person unter verschiedenen Bedingungen beobachten, verwenden Sie die Prozedur "Messwiederholungen". Diese steht Ihnen in Custom Tables und Advanced Statistics zur Verfügung.

### Berechnen einer Analyse zur Gewichtungsschätzung

1. Wählen Sie in den Menüs Folgendes 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 einer Potenz dieser Variablen gewichtet. Die Regressionsgleichung wird für jeden Wert des Exponenten in einem angegebenen Bereich berechnet und gibt den Exponenten an, der die Log-Likelihood-Funktion maximiert.
- *Exponentenbereich.* Diese Funktion wird in Verbindung mit der GewichtungsvARIABLE zur Berechnung der Gewichtungen verwendet. 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 GewichtungsvARIABLE hinzu. Diese Variable heißt *WGT<sub>n</sub>*, 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 Informationen zur Syntax finden Sie in der 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.

**Statistik.** 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 in den Menüs Folgendes aus:

**Analysieren > Regression > Zweistufige kleinste 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*. Dies sind die Variablen, die zum Berechnen der vorhergesagten Werte für die endogenen Variablen in der ersten Stufe der zweistufigen Analyse der kleinsten Quadrate verwendet werden. 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:

```
Mittelwert (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(FAKTOR)=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:

```
Mittelwert (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(FAKTOR) = 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:

```
Mittelwert (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:

$$\begin{array}{l} \text{Mittelwert ( } 1/k \quad 1/k \quad 1/k \quad \dots \quad 1/k) \\ \text{df(1) ( } -1 \quad 1 \quad 0 \quad \dots \quad 0) \\ \text{df(2) ( } -1/2 \quad -1/2 \quad 1 \quad \dots \quad 0) \\ \vdots \\ \text{df(k-1) ( } -1/(k-1) \quad -1/(k-1) \quad -1/(k-1) \quad \dots \quad 1) \end{array}$$

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

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

## 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(WIRKSTF)=POLYNOMIAL
```

dem Unterbefehl

```
/CONTRAST(WIRKSTF)=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(WIRKSTF)=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(WIRKSTF)=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 Wirkungsflächen 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:

$$\text{Mittelwert ( } 1/k \quad 1/k \quad 1/k \quad \dots \quad 1/k \quad 1/k)$$

$$\begin{array}{l} \text{df}(1) \text{ ( 1 -1 0 ... 0 0)} \\ \text{df}(2) \text{ ( 0 1 -1 ... 0 0)} \\ \vdots \\ \text{df}(k-1) \text{ ( 0 0 0 ... 1 -1)} \end{array}$$

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:

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

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)	Gewichtungen für Berechnung des Mittelwerts
(3 -1 -1 -1)	1. mit 2. bis 4. vergleichen
(0 2 -1 -1)	2. mit 3. und 4. vergleichen
(0 0 1 -1)	3. mit 4. vergleichen

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

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

Bei LOGLINEAR müssen Sie Folgendes angeben:

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

Mit Ausnahme der Zeile für den Mittelwert beträgt die Summe jeder Zeile 0. Die Summe der Produkte aller Paare in disjunkten Zeilen beträgt ebenfalls 0:

Zeilen 2 und 3:  $(3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0$   
 Zeilen 2 und 4:  $(3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0$   
 Zeilen 3 und 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  $i$ -ten Kategorie wird für fast alle Indikatorvariablen als 0 und lediglich für die  $i$ -te als 1 codiert.



---

## 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 kann eine Kopie des Produkts oder der Produktversion in der jeweiligen Sprache erforderlich sein.

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  
Tour Descartes  
2, avenue Gambetta  
92066 Paris La Defense  
France*

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 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 Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
USA*

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

Die Lieferung des in diesem Dokument beschriebenen 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 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 Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter 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 Geschäftsablaufs. 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 Programmier Techniken in verschiedenen Betriebsumgebungen veranschaulichen. Sie dürfen diese Beispielprogramme kostenlos 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:

© (Name Ihrer Firma) (Jahr). Teile des vorliegenden Codes wurden aus Beispielprogrammen der IBM Corporation abgeleitet.

© Copyright IBM Corp. \_Jahr/Jahre angeben\_. Alle Rechte vorbehalten.

---

## Marken

IBM, das IBM Logo und [ibm.com](http://ibm.com) sind Marken oder eingetragene Marken der IBM Corp in den USA und/oder anderen Ländern. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen 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 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 15

## B

- Benutzerdefinierte Modelle
  - in multinomialer logistischer Regression 7
- Binär logistische Regression 1

## C

- Chi-Quadrat nach Pearson
  - Anpassungsgüte 8
  - zum Schätzen des Skalierungswerts für Streuung 9
- 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 15

## E

- Ertragsdichtemodell
  - in nicht linearer Regression 15

## F

- Fiduzialkonfidenzintervalle
  - in Probit-Analyse 12

## G

- Gauß-Modell
  - in nicht linearer Regression 15
- Gesättigte Modelle
  - in multinomialer logistischer Regression 7
- Gewichtungsschätzung 18
  - ANOVA und Schätzungen anzeigen 19
  - Beispiel 18

- Gewichtungsschätzung (*Forts.*)
  - beste Gewichtung als neue Variable speichern 19
  - Iterationsverlauf 19
  - Log-Likelihood 18
  - Statistik 18
  - zusätzliche Funktionen beim Befehl 20
- Gompertz-Modell
  - in nicht linearer Regression 15

## H

- Hebelwerte
  - in logistischer Regression 4
- Hosmer-Lemeshow-Anpassungsstatistik
  - in logistischer Regression 5

## I

- Iteration
  - in logistischer Regression 5
  - in multinomialer logistischer Regression 9
  - in Probit-Analyse 12
- Iterationsverlauf
  - in multinomialer logistischer Regression 9

## J

- Johnson-Schumacher-Modell
  - in nicht linearer Regression 15

## K

- Kategoriale Kovariaten 4
- Klassifikation
  - in multinomialer logistischer Regression 6
- Klassifikationstabellen
  - in multinomialer logistischer Regression 8
- Konfidenzintervalle
  - in multinomialer logistischer Regression 8
- Konstanter Term
  - einschließen oder ausschließen 7
  - in linearer Regression 5
- Kontraste
  - in logistischer Regression 4
- Konvergenzkriterium
  - 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 18
  - zweistufige Regressionsmethode der kleinsten Quadrate 20
- Log-Likelihood
  - in Gewichtungsschätzung 18
  - in multinomialer logistischer Regression 8
- Log-modifiziertes Modell
  - in nicht linearer Regression 15
- Logistische Regression 2
  - 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
  - Zeichenfolgekovariaten 4
  - zusätzliche Funktionen beim Befehl 6

## M

- McFadden-R-Quadrat
  - in multinomialer logistischer Regression 8
- Metcherlich-Gesetz der abnehmenden Erträge
  - in nicht linearer Regression 15
- Michaelis-Menten-Modell
  - in nicht linearer Regression 15

- Mittlere effektive Dosis
  - in Probit-Analyse 12
- Modell mit Haupteffekten
  - in multinomialer logistischer Regression 7
- Morgan-Mercer-Florin-Modell
  - in nicht linearer Regression 15
- Multinomiale logistische Regression 6, 8
  - Exportieren von Modellinformationen 10
  - Kriterien 9
  - Modelle 7
  - Referenzkategorie 8
  - speichern 10
  - Statistik 8
  - zusätzliche Funktionen beim Befehl 11

## N

- Nagelkerke-R-Quadrat
  - in multinomialer logistischer Regression 8
- Nebenbedingungen für Parameter
  - in nicht linearer Regression 16
- Nicht lineare Modelle
  - in nicht linearer Regression 15
- Nicht lineare Regression 13
  - Ableitungen 17
  - Anfangswerte 15
  - Bedingungslogik 14
  - Beispiel 13
  - Bootstrap-Schätzungen 17
  - gängige nicht lineare Modelle 15
  - Interpretieren von Ergebnissen 17
  - Levenberg-Marquardt-Algorithmus 17
  - Nebenbedingungen für Parameter 16
  - neue Variablen speichern 17
  - Parameter 15
  - Residuen 17
  - Schätzmethoden 17
  - segmentiertes Modell 14
  - sequenzielle quadratische Optimierung 17
  - Statistik 13
  - Verlustfunktion 16
  - vorhergesagte Werte 17
  - zusätzliche Funktionen beim Befehl 18

## P

- Parallelitätstest
  - in Probit-Analyse 12
- Parameterschätzungen
  - in multinomialer logistischer Regression 8
- Peal-Reed-Modell
  - in nicht linearer Regression 15
- Probit-Analyse
  - Beispiel 11
  - Bereich definieren 12
  - Fiduzialkonfidenzintervalle 12
  - Iteration 12
  - Kriterien 12

- Probit-Analyse (*Forts.*)
  - mittlere effektive Dosis 12
  - natürliche Rücklaufquote 12
  - Parallelitätstest 12
  - Statistik 11, 12
  - zusätzliche Funktionen beim Befehl 13

## 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 16
- Richards-Modell
  - in nicht linearer Regression 15
- 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

## T

- Tabellen mit Zellenwahrscheinlichkeiten
  - in multinomialer logistischer Regression 8
- Trennung
  - in multinomialer logistischer Regression 9

## V

- Verhältnis der 3. Potenzen
  - in nicht linearer Regression 15
- Verhältnis der Quadrate
  - in nicht linearer Regression 15
- Verhulst-Modell
  - in nicht linearer Regression 15
- Von-Bertalanffy-Modell
  - in nicht linearer Regression 15
- Vorwärtsselektion
  - in logistischer Regression 3

## W

- Weibull-Modell
  - in nicht linearer Regression 15

## Z

- Zeichenfolgekovariaten
  - in logistischer Regression 4
- Zellen mit null Beobachtungen
  - in multinomialer logistischer Regression 9
- zweistufige Regressionsmethode der kleinsten Quadrate
  - Kovarianz der Parameter 21
- Zweistufige Regressionsmethode der kleinsten Quadrate 20
  - Beispiel 20
  - Instrumentalvariablen 20
  - Speichern von neuen Variablen 21
  - Statistik 20
  - zusätzliche Funktionen beim Befehl 21



