

*IBM SPSS Advanced Statistics 29*

**IBM**

**Hinweis**

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 147 gelesen werden.

**Produktinformation**

Diese Ausgabe bezieht sich auf Version 29, Release 0, Modifikation 2 von IBM® SPSS Statistics und auf alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuausgabe geändert wird.

© **Copyright International Business Machines Corporation .**

---

# Inhaltsverzeichnis

<b>Kapitel 1. Erweiterte Statistiken.....</b>	<b>1</b>
Einführung in Advanced Statistics.....	1
GLM - Multivariat.....	1
GLM - Multivariat: Modell.....	3
Multivariat: Kontraste.....	4
GLM - Multivariat: Profilplots.....	5
GLM - Multivariat: Post-hoc-Vergleiche.....	5
GLM: Geschätzte Randmittel.....	6
GLM: Speichern.....	7
Multivariat: Optionen.....	8
Zusätzliche Funktionen beim Befehl GLM.....	9
GLM - Messwiederholungen.....	10
Messwiederholungen: Faktor(en) definieren.....	12
Messwiederholungen: Modell.....	12
Messwiederholungen: Kontraste.....	14
GLM - Messwiederholung: Profilplots.....	14
GLM - Messwiederholung: Post-hoc-Mehrfachvergleiche.....	15
GLM: Geschätzte Randmittel.....	16
GLM - Messwiederholungen: Speichern.....	17
GLM - Messwiederholungen: Optionen.....	17
Zusätzliche Funktionen beim Befehl GLM.....	18
Analyse der Varianzkomponenten.....	19
Varianzkomponenten: Modell.....	20
Varianzkomponenten: Optionen.....	21
Varianzkomponenten: In neuer Datei speichern.....	22
Zusätzliche Funktionen beim Befehl VARCOMP.....	22
Lineare gemischte Modelle.....	22
Lineare gemischte Modelle: Subjekte und Wiederholungen.....	24
Lineare gemischte Modelle: Feste Effekte.....	25
Lineare gemischte Modelle: Zufällige Effekte.....	26
Lineare gemischte Modelle: Schätzung.....	27
Lineare gemischte Modelle: Statistik.....	28
Lineare gemischte Modelle: Geschätzte Randmittel.....	29
Lineare gemischte Modelle: Speichern.....	29
Lineare gemischte Modelle-Exportieren.....	30
Zusätzliche Funktionen beim Befehl MIXED.....	30
Verallgemeinerte lineare Modelle.....	30
Verallgemeinerte lineare Modelle: Antwort.....	33
Verallgemeinerte lineare Modelle: Prädiktoren.....	34
Verallgemeinerte lineare Modelle: Modell.....	34
Verallgemeinerte lineare Modelle: Schätzung.....	35
Verallgemeinerte lineare Modelle: Statistiken.....	37
Verallgemeinerte lineare Modelle: Geschätzte Randmittel.....	38
Verallgemeinerte lineare Modelle: Speichern.....	39
Verallgemeinerte lineare Modelle: Exportieren.....	40
Zusätzliche Funktionen beim Befehl GENLIN.....	41
Verallgemeinerte Schätzungsgleichungen.....	41
Verallgemeinerte Schätzungsgleichungen: Typ des Modells.....	43
Verallgemeinerte Schätzungsgleichungen: Antwort.....	45
Verallgemeinerte Schätzungsgleichungen: Prädiktoren.....	46
Verallgemeinerte Schätzungsgleichungen: Modell.....	47

Verallgemeinerte Schätzungsgleichungen: Schätzung.....	47
Verallgemeinerte Schätzungsgleichungen: Statistiken.....	49
Verallgemeinerte Schätzungsgleichungen: Geschätzte Randmittel.....	50
Verallgemeinerte Schätzungsgleichungen: Speichern.....	51
Verallgemeinerte Schätzungsgleichungen: Exportieren.....	52
Zusätzliche Funktionen beim Befehl GENLIN.....	53
Verallgemeinerte lineare gemischte Modelle.....	53
Erzeugen eines verallgemeinerten linearen gemischten Modells.....	55
Ziel .....	56
Feste Effekte .....	58
Zufällige Effekte .....	59
Gewichtung und Offset .....	61
Allgemeine Erstellungsoptionen .....	61
Schätzung.....	62
Geschätzte Mittelwerte .....	62
Speichern .....	63
Exportieren .....	64
Modellansicht .....	64
Modellauswahl für loglineare Analyse.....	68
Loglineare Analyse: Bereich definieren.....	69
Loglineare Analyse: Modell.....	69
Modellauswahl für loglineare Analyse: Optionen.....	70
Zusätzliche Funktionen beim Befehl HILOGLINEAR.....	70
Allgemeine loglineare Analyse.....	70
Allgemeine loglineare Analyse: Modell.....	72
Allgemeine loglineare Analyse: Optionen.....	72
Allgemeine loglineare Analyse: Speichern.....	73
Zusätzliche Funktionen beim Befehl GENLOG.....	73
Logit-loglineare Analyse.....	73
Logit-loglineare Analyse: Modell.....	74
Logit-loglineare Analyse: Optionen.....	75
Logit-loglineare Analyse: Speichern.....	76
Zusätzliche Funktionen beim Befehl GENLOG.....	76
Sterbetafel.....	76
Sterbetafel: Bereich für Faktorvariable definieren.....	78
Sterbetafel: Optionen.....	78
Zusätzliche Funktionen beim Befehl SURVIVAL.....	78
Parametrische beschleunigte Fehlerzeitmodelle.....	78
Parametrische beschleunigte Fehlerzeitmodelle: Kriterien.....	79
Modelle für parametrische beschleunigte Ausfallzeit: Modell.....	80
Parametrische beschleunigte Fehlerzeitmodelle: Schätzung.....	81
Parametrische beschleunigte Fehlerzeitmodelle: Drucken.....	82
Parametrische Modelle für beschleunigte Ausfallzeiten: Vorhersagen.....	83
Modelle für parametrische beschleunigte Ausfallzeit: Diagramm.....	83
Parametrische beschleunigte Fehlerzeitmodelle: Exportieren.....	84
Survival AFT Ereignisse für Statusvariablen definieren.....	84
Parametrische beschleunigte Fehlerzeitmodelle: Kategorie auswählen.....	84
Parametrische gemeinsam genutzte Frailty-Modelle.....	85
Parametrische gemeinsam genutzte Frailty-Modelle: Kriterien.....	86
Parametrische gemeinsam genutzte Frailty-Modelle: Modell.....	87
Parametrische gemeinsam genutzte Frailty-Modelle: Schätzung.....	88
Parametrische gemeinsam genutzte Frailty-Modelle: Drucken.....	89
Parametrische gemeinsam genutzte Frailty-Modelle: Vorhersagen.....	90
Parametrische gemeinsam genutzte Frailty-Modelle: Diagramm.....	91
Parametrische gemeinsam genutzte Frailty-Modelle: Exportieren.....	91
Parametrische gemeinsam genutzte Frailty-Modelle: Ereignisse definieren.....	91
Parametrische gemeinsam genutzte Frailty-Modelle-Beispiele.....	92
Parametrische gemeinsam genutzte Frailty-Modelle-Fallstudie für Rekurrent-Daten.....	94

Überlebensanalyse nach Kaplan-Meier.....	96
Kaplan-Meier: Ereignis für Statusvariable definieren.....	97
Kaplan-Meier: Faktorstufen vergleichen.....	97
Kaplan-Meier: Neue Variablen speichern.....	98
Kaplan-Meier: Optionen.....	98
Zusätzliche Funktionen beim Befehl KM.....	98
Cox-Regressionsanalyse.....	99
Cox-Regression: Kategoriale Kovariaten definieren.....	99
Cox-Regression: Diagramme.....	100
Cox-Regression: Neue Variablen speichern.....	101
Cox-Regression: Optionen.....	101
Cox-Regression: Ereignis für Statusvariable definieren.....	102
Zusätzliche Funktionen beim Befehl COXREG.....	102
Berechnen von zeitabhängigen Kovariaten.....	102
Berechnen einer zeitabhängigen Kovariate.....	103
Codierungsschemata für kategoriale Variablen.....	103
Abweichung.....	103
Einfach.....	104
Helmert.....	104
Differenz.....	104
Polynomial.....	105
Wiederholt.....	105
Spezial.....	106
Indikator.....	106
Kovarianzstrukturen.....	107
Bayessche Statistik.....	111
Bayes-Inferenz bei einer Stichprobe: Normal.....	112
Bayes-Inferenz bei einer Stichprobe: Binomial.....	116
Bayes-Inferenz bei einer Stichprobe: Poisson.....	118
Bayes-Inferenz bei verbundenen Stichproben: Normal.....	120
Bayes-Inferenz bei unabhängigen Stichproben.....	122
Bayessche Inferenz über Pearson-Korrelation.....	126
Bayessche Inferenz über lineare Regressionsmodelle.....	128
Bayessche einfaktorielle ANOVA.....	134
Bayessche loglineare Modelle.....	137
Bayessche einfaktorielle ANOVA mit Messwiederholung: Modelle.....	141
Kernel-Ridge-Regression.....	143
Kernelparameter.....	144
Kernel-Ridge-Regression: Optionen.....	144
<b>Bemerkungen.....</b>	<b>147</b>
Marken.....	148
<b>Index.....</b>	<b>151</b>



---

# Kapitel 1. Erweiterte Statistiken

Die folgenden Funktionen für erweiterte Statistiken sind in SPSS Statistics Standard Edition oder der Option "Advanced Statistics" enthalten.

---

## Einführung in Advanced Statistics

Die Option "Erweiterte Statistik" enthält Prozeduren, die erweiterte Modellierungsoptionen anbieten, als über die Option "Statistikbasis" verfügbar sind.

- "GLM - Multivariat" erweitert das von "GLM - Univariat" bereitgestellte allgemeine lineare Modell dahingehend, dass mehrere abhängige Variablen zulässig sind. Eine andere Erweiterung, "GLM - Messwiederholung" ermöglicht wiederholte Messungen mehrerer abhängiger Variablen.
- "Analyse von Varianzkomponenten" ist ein spezielles Tool zur Zerlegung der Variabilität in einer abhängigen Variablen in feste Komponenten und Zufallskomponenten.
- Die Prozedur "Lineare gemischte Modelle" erweitert das allgemeine lineare Modell, indem sie zulässt, dass die Daten korrelierte und nicht konstante Variabilität aufweisen können. Das gemischte lineare Modell gibt Ihnen daher die Flexibilität, nicht nur die Mittelwerte der Daten, sondern auch die Varianzen und Kovarianzen zu analysieren.
- "Verallgemeinerte Lineare Modelle" (GZLM) lockert die Voraussetzung der Normalität für den Fehlerterm und erfordert nur, dass die abhängige Variable über eine Transformation oder Verknüpfungsfunktion in einem linearen Zusammenhang mit den Prädiktoren steht. "Verallgemeinerte Schätzungsgleichungen" (GEE) erweitert GLZM dahingehend, dass wiederholte Messungen zulässig sind.
- Die Funktion "Allgemeine loglineare Analyse" ermöglicht die Anpassung von Modellen für kreuzklassifizierte Häufigkeitsdaten und die Modellauswahl für loglineare Analyse ermöglicht Ihnen die Auswahl zwischen verschiedenen Modellen.
- Mit "Logit-loglineare Analyse" können Sie loglineare Modelle anpassen, um die Beziehung zwischen einer kategorialen abhängigen Variablen und einer oder mehreren kategorialen Prädiktoren zu analysieren.
- Eine Überlebensanalyse ist mithilfe von Sterbetafeln möglich, mit denen die Verteilung von Variablen für die Zeit bis zum Eintreten des Ereignisses untersucht werden kann, möglicherweise anhand der Stufen einer Faktorvariablen, Überlebensanalyse nach Kaplan-Meier zur Untersuchung der Verteilung von Variablen für die Zeit bis zum Eintreten des Ereignisses, möglicherweise anhand der Stufen einer Faktorvariablen oder unter Erstellung von separaten Analysen anhand der Stufen einer Schichtungsvariablen; sowie die Cox-Regression zur Modellierung der Zeit bis zu einem angegebenen Ereignis, auf der Grundlage der Werte der jeweils vorgegebenen Kovariaten.

---

## GLM - Multivariat

Mit der Prozedur "GLM - Multivariat" können Sie Regressionsanalysen und Varianzanalysen für mehrere abhängige Variablen mit einer oder mehreren Faktorvariablen oder Kovariaten vornehmen. Die Faktorvariablen unterteilen die Grundgesamtheit in Gruppen. Mit dieser Prozedur, die von einem allgemeinen linearen Modell ausgeht, können Sie Nullhypothesen über die Auswirkungen von Faktorvariablen auf die Mittelwerte verschiedener Gruppierungen einer gemeinsamen Verteilung von abhängigen Variablen testen. Sie können sowohl die Interaktionen zwischen den Faktoren als auch die Auswirkungen einzelner Faktoren untersuchen. Außerdem können Sie die Auswirkungen von Kovariaten und Interaktionen zwischen Kovariaten und Faktoren berücksichtigen. Bei der Regressionsanalyse werden die unabhängigen Variablen (Prädiktorvariablen) als Kovariaten angegeben.

Es können sowohl ausgewogene als auch unausgewogene Modelle getestet werden. Ein Design ist ausgewogen, wenn jede Zelle im Modell dieselbe Anzahl von Fällen enthält. In einem multivariaten Modell liegen die Quadratsummen aus den Effekten im Modell und die Fehlerquadratsummen in Matrixform und nicht wie bei der univariaten Analyse in skalarer Form vor. Diese Matrizen werden als SSCP-Matrizen

(Matrizen der Quadratsummen und Kreuzprodukte) bezeichnet. Wenn Sie mehr als eine abhängige Variable angegeben haben, werden sowohl die multivariate Varianzanalyse unter Verwendung der Pillai-Spur, Wilks-Lambda, Hotelling-Spur und Roys größter charakteristischer Wurzel mit angenäherter  $F$ -Statistik als auch die univariate Varianzanalyse für jede abhängige Variable ausgegeben. Mit der Prozedur "GLM - Multivariat" werden nicht nur Hypothesen getestet, sondern zugleich Parameter geschätzt.

Zum Testen von Hypothesen stehen häufig verwendete *A-priori*-Kontraste zur Verfügung. Nachdem die Signifikanz mit einem  $F$ -Gesamttest nachgewiesen wurde, können Sie Post-hoc-Tests verwenden, um Differenzen zwischen bestimmten Mittelwerten berechnen zu lassen. Geschätzte Randmittel dienen als Schätzungen für die vorhergesagten Mittelwerte der Zellen im Modell und mit Profilplots (Interaktionsdiagrammen) dieser Mittelwerte können Sie einige dieser Beziehungen in einfacher Weise visuell darstellen. Die Post-hoc-Mehrfachvergleiche werden für jede abhängige Variable getrennt durchgeführt.

Residuen, Einflusswerte, die Cook-Distanz und Hebelwerte können zum Überprüfen von Annahmen als neue Variablen in der Datendatei gespeichert werden. Weiterhin stehen eine SSCP-Matrix für Residuen (eine quadratische Matrix von Quadratsummen und Kreuzprodukten für Residuen), eine Residuenkovarianzmatrix (die SSCP-Matrix für Residuen geteilt durch die Freiheitsgrade der Residuen) und die Residuenkorrelationsmatrix (die standardisierte Form der Residuenkovarianzmatrix) zur Verfügung.

Mit der WLS-Gewichtung können Sie eine Variable angeben, um Beobachtungen für eine WLS-Analyse (Weighted Least Squares - gewichtete kleinste Quadrate) unterschiedlich zu gewichten. Dies kann notwendig sein, um etwaige Unterschiede in der Präzision von Messungen auszugleichen.

**Beispiel.** Ein Hersteller von Kunststoffen misst drei Eigenschaften von Plastikfolie: Reißfestigkeit, Glanz und Lichtundurchlässigkeit. Es werden zwei Produktionsgeschwindigkeiten und zwei verschiedene Anteile von Zusätzen ausprobiert und die drei Eigenschaften werden bei jeder Kombination dieser Bedingungen gemessen. Es stellt sich heraus, dass die Produktionsgeschwindigkeit und der Anteil an Zusätzen jeweils signifikante Ergebnisse ergeben, dass aber die Interaktion der beiden Faktoren nicht signifikant ist.

**Methoden.** Zum Überprüfen der verschiedenen Hypothesen können Quadratsummen vom Typ I, Typ II, Typ III und Typ IV verwendet werden. Die Voreinstellung sieht den Typ III vor.

**Statistik** Post-hoc-Spannweitentests und Mehrfachvergleiche: geringste signifikante Differenz, Bonferoni, Sidak, Scheffé, multiples  $F$  nach Ryan-Einot-Gabriel-Welsch, multiple Spannweite nach Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls-Test, Tukey-HSD-Test, Tukey- $B$ , Duncan, GT2 nach Hochberg, Gabriel, Waller-Duncan- $T$ -Test, Dunnett (einseitig und zweiseitig), Tamhane- $T_2$ , Dunnett- $T_3$ , Games-Howell und Dunnett- $C$ . Deskriptive Statistiken: beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten aller abhängigen Variablen in allen Zellen, Levene-Test auf Homogenität der Varianzen, Box- $M$ -Test auf Homogenität der Kovarianzmatrizen für die abhängigen Variablen und Bartlett-Test auf Sphärität.

**Diagramme.** Streubreite gegen mittleres Niveau, Residuendiagramme, Profilplots (Interaktion).

Erläuterungen der Daten für GLM - Multivariat

**Daten.** Die abhängigen Variablen müssen quantitativ sein. Faktoren sind kategorial und können numerische Werte oder Zeichenfolgewerte besitzen. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen.

**Annahmen.** Die Daten für die abhängigen Variablen werden aus einer Zufallsstichprobe von Vektoren aus einer multivariaten, normalverteilten Grundgesamtheit entnommen. In der Grundgesamtheit sind die Varianz-Kovarianz-Matrizen für alle Zellen gleich. Die Varianzanalyse ist unempfindlich gegenüber Abweichungen von der Normalverteilung. Die Daten müssen jedoch symmetrisch verteilt sein. Zum Überprüfen der Annahmen können Sie Tests auf Homogenität der Varianzen vornehmen (einschließlich Box- $M$ ) und Diagramme der Streubreite gegen das mittlere Niveau anfordern. Sie können auch die Residuen untersuchen und Residuendiagramme anzeigen lassen.

**Verwandte Prozeduren.** Verwenden Sie die Prozedur "Explorative Datenanalyse" zum Untersuchen der Daten, ehe Sie eine Varianzanalyse durchführen. Verwenden Sie bei einer einzelnen abhängigen Variablen die Prozedur "GLM - Univariat". Wenn Sie dieselben abhängigen Variablen zu mehreren Zeitpunkten für jedes Subjekt gemessen haben, verwenden Sie die Prozedur "GLM - Messwiederholungen".

Berechnen einer multivariaten Analyse der Varianz (GLM)

1. Wählen Sie Folgendes in den Menüs aus:

## Analysieren > Allgemeines lineares Modell > Multivariat...

2. Wählen Sie mindestens zwei abhängige Variablen aus.

Wahlweise können Sie feste Faktoren, Kovariaten und eine WLS-Gewichtung angeben.

## GLM - Multivariat: Modell

**Modell angeben.** Ein gesättigtes Modell enthält alle Haupteffekte der Faktoren, alle Kovariatenhaupteffekte und alle Interaktionen zwischen Faktoren. Es enthält keine Kovariateninteraktionen. Wählen Sie **Anpassen** aus, um nur ein Subset von Interaktionen oder Interaktionen zwischen Faktoren und Kovariaten festzulegen. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

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

**Modell.** Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von **Anpassen** können Sie die Haupteffekte und Interaktionen auswählen, die für Ihre Analyse von Interesse sind.

**Quadratsumme.** Hier wird die Methode zum Berechnen der Quadratsumme festgelegt. Für ausgewogene und unausgewogene Modelle ohne fehlende Zellen wird meistens die Methode mit Quadratsummen vom Typ III angewendet.

**Konstanter Term in Modell einschließen.** Der konstante Term wird gewöhnlich in das Modell aufgenommen. Falls Sie sicher sind, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

## Erstellen von Termen und benutzerdefinierten Termen

### Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie nicht verschachtelte Terme eines bestimmten Typs (beispielsweise Haupteffekte) für alle Kombinationen eines ausgewählten Sets von Faktoren und Kovariaten einschließen wollen.

### Benutzerdefinierte Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie verschachtelte Terme einschließen oder einen Term explizit Variable für Variable erstellen wollen. Die Erstellung eines verschachtelten Terms besteht aus folgenden Schritten:

## Quadratsumme

Für das Modell können Sie einen Typ von Quadratsumme auswählen. Typ III wird am häufigsten verwendet und ist die Standardeinstellung.

**Typ I.** Diese Methode ist auch als die Methode der hierarchischen Zerlegung der Quadratsummen bekannt. Jeder Term wird nur für den Vorläuferterm im Modell korrigiert. Quadratsummen vom Typ I werden gewöhnlich in den folgenden Situationen verwendet:

- Ein ausgewogenes ANOVA-Modell, in dem alle Haupteffekte vor den Interaktionseffekten 1. Ordnung festgelegt werden, alle Interaktionseffekte 1. Ordnung wiederum vor den Interaktionseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme niedrigerer Ordnung vor den Termen höherer Ordnung festgelegt werden.
- Ein rein verschachteltes Modell, in welchem der zuerst bestimmte Effekt in dem als zweiten bestimmten Effekt verschachtelt ist, der zweite Effekt wiederum im dritten und so weiter. (Diese Form der Verschachtelung kann nur durch Verwendung der Befehlssprache erreicht werden.)

**Typ II.** Bei dieser Methode wird die Quadratsumme eines Effekts im Modell angepasst an alle anderen "zutreffenden" Effekte berechnet. Ein zutreffender Effekt ist ein Effekt, der mit allen Effekten in Beziehung steht, die den untersuchten Effekt nicht enthalten. Die Methode mit Quadratsummen vom Typ II wird gewöhnlich in den folgenden Fällen verwendet:

- Bei ausgewogenen ANOVA-Modellen.
- Bei Modellen, die nur Haupteffekte von Faktoren enthalten.

- Jedes Regressionsmodell.
- Bei rein verschachtelten Designs. (Diese Form der Verschachtelung kann durch Verwendung der Befehlssprache erreicht werden.)

**Typ III.** Der Standardparameter. Bei dieser Methode werden die Quadratsummen eines Effekts im Design als Quadratsummen orthogonal zu allen Effekten (sofern vorhanden), die den Effekt enthalten, und mit Bereinigung um alle anderen Effekte berechnet, die den Effekt nicht enthalten. Der große Vorteil der Quadratsummen vom Typ III ist, dass sie invariant bezüglich der Zellenhäufigkeiten sind, solange die allgemeine Form der Schätzbarkeit konstant bleibt. Daher wird dieser Typ von Quadratsumme oft für unausgewogene Modelle ohne fehlende Zellen als geeignet angesehen. In einem faktoriellen Design ohne fehlende Zellen ist diese Methode äquivalent zu der Methode der gewichteten Mittelwertquadrate nach Yates. Die Methode mit Quadratsummen vom Typ III wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgewogenen oder unausgewogenen Modelle ohne leere Zellen.

**Typ IV.** Diese Methode ist dann geeignet, wenn es keine fehlenden Zellen gibt. Für alle Effekte  $F$  im Design: Wenn  $F$  in keinem anderen Effekt enthalten ist, dann gilt: Typ IV = Typ III = Typ II. Wenn  $F$  in anderen Effekten enthalten ist, werden bei Typ IV die Kontraste zwischen den Parametern in  $F$  gleichmäßig auf alle Effekte höherer Ordnung verteilt. Die Methode mit Quadratsummen vom Typ IV wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgewogenen oder unausgewogenen Modelle mit leeren Zellen.

## Multivariat: Kontraste

Kontraste werden verwendet, um zu testen, ob sich die Faktorstufen signifikant voneinander unterscheiden. Sie können für jeden Faktor im Modell einen Kontrast festlegen. Kontraste stellen lineare Kombinationen der Parameter dar.

Das Testen der Hypothesen basiert auf der Nullhypothese  $\mathbf{LBM} = \mathbf{0}$ . Dabei ist  $\mathbf{L}$  die L-Matrix der Kontrastkoeffizienten,  $\mathbf{M}$  die Identitätsmatrix (deren Dimension gleich der Anzahl der abhängigen Variablen ist) und  $\mathbf{B}$  ist der Parametervektor. Wenn Sie einen Kontrast festlegen, wird eine  $\mathbf{L}$ -Matrix erstellt, bei der die Spalten für den betreffenden Faktor mit dem Kontrast übereinstimmen. Die verbleibenden Spalten werden so angepasst, dass die  $\mathbf{L}$ -Matrix schätzbar ist.

Zusätzlich zu den univariaten Tests unter Verwendung der  $F$ -Statistik und den simultanen Konfidenzintervallen nach Bonferroni auf der Grundlage der Student- $T$ -Verteilung für die Kontrastdifferenzen zwischen allen abhängigen Variablen wird der multivariate Test unter Verwendung der Pillai-Spur, Wilk-Lambda, Hotelling-Spur und Roys größter charakteristischer Wurzel ausgegeben.

Als Kontraste sind "Abweichung", "Einfach", "Differenz", "Helmert", "Wiederholt" und "Polynomial" verfügbar. Bei Abweichungskontrasten und einfachen Kontrasten können Sie wählen, ob die letzte oder die erste Kategorie als Referenzkategorie dient.

### Kontrasttypen

**Abweichung.** Vergleicht den Mittelwert jeder Faktorstufe (außer bei Referenzkategorien) mit dem Mittelwert aller Faktorstufen (Gesamtmittelwert). Die Stufen des Faktors können in beliebiger Ordnung vorliegen.

**Einfach.** Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt. Sie können die erste oder die letzte Kategorie als Referenz auswählen.

**Differenz.** Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. (Dies wird gelegentlich auch als umgekehrter Helmert-Kontrast bezeichnet).

**Helmert.** Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.

**Wiederholt.** Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.

**Polynomial.** Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

## GLM - Multivariat: Profilplots

Profilplots (Interaktionsdiagramme) sind hilfreich zum Vergleichen von Randmitteln im Modell. Ein Profilplot ist ein Liniendiagramm, in dem jeder Punkt das geschätzte Randmittel einer abhängigen Variablen (angepasst an die Kovariaten) bei einer Stufe eines Faktors angibt. Die Stufen eines zweiten Faktors können zum Erzeugen getrennter Linien verwendet werden. Jede Stufe in einem dritten Faktor kann verwendet werden, um ein separates Diagramm zu erzeugen. Für Plots stehen alle Faktoren zur Verfügung. Profilplots werden für jede einzelne abhängige Variable erstellt.

Ein Profilplot für einen Faktor zeigt, ob die geschätzten Randmittel mit den Faktorstufen steigen oder fallen. Bei zwei oder mehr Faktoren deuten parallele Linien an, dass es keine Interaktion zwischen den Faktoren gibt. Das heißt, dass Sie die Faktorstufen eines einzelnen Faktors untersuchen können. Nicht parallele Linien deuten auf eine Interaktion hin.

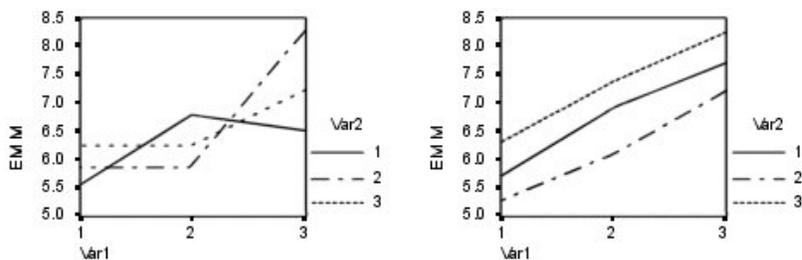


Abbildung 1. Nicht paralleles Diagramm (links) und paralleles Diagramm (rechts)

Nachdem ein Diagramm durch Auswahl von Faktoren für die horizontale Achse (und wahlweise von Faktoren für getrennte Linien und getrennte Diagramme) festgelegt wurde, muss das Diagramm der Liste "Diagramme" hinzugefügt werden.

## GLM - Multivariat: Post-hoc-Vergleiche

**Tests für Post-hoc-Mehrfachvergleiche.** Sobald Sie festgestellt haben, dass es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-hoc-Spannweitentests und paarweisen multiplen Vergleichen untersuchen, welche Mittelwerte sich unterscheiden. Die Vergleiche werden auf der Basis von nicht korrigierten Werten vorgenommen. Die Post-hoc-Tests werden für jede abhängige Variable getrennt durchgeführt.

Häufig verwendete Mehrfachvergleiche sind der Bonferroni-Test und der Tukey-HSD-Test. Der **Bonferroni-Test** auf der Grundlage der studentisierten  $T$ -Statistik passt das beobachtete Signifikanzniveau unter Berücksichtigung der Tatsache an, dass multiple Vergleiche vorgenommen werden. Der **Sidak-t-Test** korrigiert ebenfalls das Signifikanzniveau und liefert engere Grenzen als der Bonferroni-Test. Der **Tukey-HSD-Test** verwendet die studentisierte Spannweitenstatistik, um alle paarweisen Vergleiche zwischen den Gruppen vorzunehmen, und setzt die experimentelle Fehlerrate auf die Fehlerrate der Ermittlung aller paarweisen Vergleiche. Beim Testen einer großen Anzahl von Mittelwertpaaren ist der Tukey-HSD-Test leistungsfähiger als der Bonferroni-Test. Bei einer kleinen Anzahl von Paaren ist der Bonferroni-Test leistungsfähiger.

**GT2 nach Hochberg** ähnelt dem Tukey-HSD-Test, es wird jedoch der studentisierte Maximalmodulus verwendet. Meistens ist der Tukey-HSD-Test leistungsfähiger. Der **paarweise Vergleichstest nach Gabriel** verwendet ebenfalls den studentisierten Maximalmodulus und zeigt meistens eine größere Schärfe als das GT2 nach Hochberg, wenn die Zellengrößen ungleich sind. Der Gabriel-Test kann ungenau werden, wenn die Zellengrößen stark variieren.

Mit dem **paarweisen t-Test für mehrere Vergleiche nach Dunnett** wird ein Set von Verarbeitungen mit einem einzelnen Kontrollmittelwert verglichen. Als Kontrollkategorie ist die letzte Kategorie voreingestellt. Sie können aber auch die erste Kategorie einstellen. Außerdem können Sie einen einseitigen oder zweiseitigen Test wählen. Verwenden Sie einen zweiseitigen Test, um zu überprüfen, ob sich der Mittelwert bei jeder Stufe (außer der Kontrollkategorie) des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet. Um zu überprüfen, ob sich der Mittelwert bei jeder Stufe des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet, wählen Sie **< Control**. Wählen Sie **>Steuerung** aus, um zu testen, ob der Mittelwert auf einer beliebigen Ebene des Faktors größer ist als die der Kontrollkategorie.

Ryan, Einot, Gabriel und Welsch (R-E-G-W) entwickelten zwei multiple Step-down-Spannweitentests. Multiple Step-down-Prozeduren überprüfen zuerst, ob alle Mittelwerte gleich sind. Wenn nicht alle Mittelwerte gleich sind, werden Subsets der Mittelwerte auf Gleichheit getestet. Das **F nach R-E-G-W** basiert auf einem *F*-Test und **Q nach R-E-G-W** basiert auf der studentisierten Spannweite. Diese Tests sind leistungsfähiger als der multiple Spannweitentest nach Duncan und der Student-Newman-Keuls-Test (ebenfalls multiple Step-down-Prozeduren), aber sie sind bei ungleichen Zellengrößen nicht empfehlenswert.

Bei ungleichen Varianzen verwenden Sie **Tamhane-T2** (konservativer paarweiser Vergleichstest auf der Grundlage eines *T*-Tests), **Dunnett-T3** (paarweiser Vergleichstest auf der Grundlage des studentisierten Maximalmodulus), den **paarweisen Vergleichstest nach Games-Howell** (manchmal ungenau) oder das **Dunnett-C** (paarweiser Vergleichstest auf der Grundlage der studentisierten Spannweite).

Der **multiple Spannweitentest nach Duncan**, Student-Newman-Keuls (**S-N-K**) und **Tukey-B** sind Spannweitentests, mit denen Gruppenmittelwerte geordnet und ein Wertebereich berechnet wird. Diese Tests werden nicht so häufig verwendet wie die vorher beschriebenen Tests.

Der **Waller-Duncan-t-Test** verwendet die Bayes-Methode. Dieser Spannweitentest verwendet den harmonischen Mittelwert der Stichprobengröße, wenn die Stichprobengrößen ungleich sind.

Das Signifikanzniveau des **Scheffé**-Tests ist so festgelegt, dass alle möglichen linearen Kombinationen von Gruppenmittelwerten getestet werden können und nicht nur paarweise Vergleiche verfügbar sind, wie bei dieser Funktion der Fall. Das führt dazu, dass der Scheffé-Test oftmals konservativer als andere Tests ist, also für eine Signifikanz eine größere Differenz der Mittelwerte erforderlich ist.

Der paarweise multiple Vergleichstest auf geringste signifikante Differenz (**LSD**) ist äquivalent zu multiplen individuellen *T*-Tests zwischen allen Gruppenpaaren. Der Nachteil bei diesem Test ist, dass kein Versuch unternommen wird, das beobachtete Signifikanzniveau im Hinblick auf multiple Vergleiche zu korrigieren.

**Angezeigte Tests.** Es werden paarweise Vergleiche für LSD, Sidak, Bonferroni, Games-Howell, T2 und T3 nach Tamhane, Dunnett-C und Dunnett-T3 ausgegeben. Homogene Subsets für Spannweitentests werden ausgegeben für S-N-K, Tukey-B, Duncan, *F* nach R-E-G-W, *Q* nach R-E-G-W und Waller. Der Tukey-HSD-Test, der GT2 nach Hochberg, der Gabriel-Test und der Scheff-Test sind multiple Vergleiche, zugleich aber auch Spannweitentests.

## GLM: Geschätzte Randmittel

Wählen Sie die Faktoren und Interaktionen aus, für die Sie Schätzungen für die Randmittel der Grundgesamtheit in den Zellen wünschen. Diese Mittel werden gegebenenfalls an die Kovariaten angepasst.

### Haupteffekte vergleichen

Gibt nicht korrigierte paarweise Vergleiche zwischen den geschätzten Randmitteln für alle Haupteffekte im Modell aus, sowohl für Zwischensubjektfaktoren als auch für Innersubjektfaktoren. Diese Option ist nur verfügbar, falls in der Liste "Mittelwerte anzeigen für" Haupteffekte ausgewählt sind.

### Einfache Haupteffekte vergleichen

Diese Einstellung wird immer dann aktiviert, wenn die Zielliste mindestens einen Produkt- oder Interaktionseffekt enthält (zum Beispiel  $A*B$ ,  $A*B*C$ ). Die Einstellung unterstützt die Spezifikation von Vergleichen bei einfachen Haupteffekten, bei denen es sich um Haupteffekte handelt, die innerhalb der Stufen anderer Faktoren verschachtelt sind.

### Anpassung des Konfidenzintervalls

Wählen Sie für das Konfidenzintervall und die Signifikanz entweder die geringste signifikante Differenz (LSD - Least Significant Difference), Bonferroni oder die Anpassung nach Sidak. Diese Option ist nur verfügbar, wenn **Haupteffekte vergleichen** und/oder **Einfache Haupteffekte vergleichen** ausgewählt ist.

### Angeben geschätzter Randmittel

1. Wählen Sie aus den Menüs eine der unter > **Analysieren** > **Allgemeines lineares Modell** verfügbaren Prozeduren aus.
2. Klicken Sie im Hauptdialogfeld auf **Geschätzte Randmittel**.

## GLM: Speichern

Vom Modell vorhergesagte Werte, Residuen und verwandte Maße können als neue Variablen im Dateneditor gespeichert werden. Viele dieser Variablen können zum Untersuchen von Annahmen über die Daten verwendet werden. Um die Werte zur Verwendung in einer anderen IBM SPSS Statistics-Sitzung zu speichern, müssen Sie die aktuelle Datendatei speichern.

**Vorhergesagte Werte.** Dies sind die Werte, welche das Modell für die einzelnen Fälle vorhersagt.

- *Nicht standardisiert.* Der Wert, den das Modell für die abhängige Variable vorhersagt.
- *Gewichtet.* Gewichtete nicht standardisierte vorhergesagte Werte. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- *Standardfehler.* Ein Schätzwert der Standardabweichung des Durchschnittswertes der abhängigen Variablen für die Fälle, die dieselben Werte für die unabhängigen Variablen haben.

**Diagnose.** Dies sind Maße zum Auffinden von Fällen mit ungewöhnlichen Wertekombinationen bei den unabhängigen Variablen und von Fällen, die einen großen Einfluss auf das Modell haben könnten.

- *Cook-Distanz.* Ein Maß dafür, wie sehr sich die Residuen aller Fälle ändern würden, wenn ein bestimmter Fall aus der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, dass der Ausschluss eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substantiell verändert.
- *Hebelwerte.* Nicht zentrierte Hebelwerte. Der relative Einfluss einer jeden Beobachtung auf die Anpassungsgüte eines Modells.

**Residuen.** Ein nicht standardisiertes Residuum ist der tatsächliche Wert der abhängigen Variablen minus des vom Modell geschätzten Werts. Ebenfalls verfügbar sind standardisierte, studentisierte und ausgeschlossene Residuen. Falls Sie eine WLS-Variable ausgewählt haben, sind auch gewichtete nicht standardisierte Residuen verfügbar.

- *Nicht standardisiert.* Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- *Gewichtet.* Gewichtete nicht standardisierte Residuen. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- *Standardisiert.* Das Residuum dividiert durch eine Schätzung der Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- *Studentisiert.* Das Residuum dividiert durch eine Schätzung seiner Standardabweichung, die von Fall zu Fall variiert, abhängig von der Entfernung der Werte jedes Falls auf den unabhängigen Variablen vom Mittelwert der unabhängigen Variablen. Wird manchmal als intern studentisiertes Residuum bezeichnet.
- *Gelöscht.* Das Residuum für einen Fall, wenn dieser Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen wird. Dies ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.

**Koeffizientenstatistik.** Hiermit wird eine Varianz-Kovarianz-Matrix der Parameterschätzungen für das Modell in ein neues Dataset in der aktuellen Sitzung oder in eine externe Datei im IBM SPSS Statistics-

Format geschrieben. Für jede abhängige Variable gibt es außerdem eine Zeile mit Parameterschätzungen, eine Zeile mit Standardfehlern der Parameterschätzungen, eine Zeile mit Signifikanzwerten für die *t*-Statistik der betreffenden Parameterschätzungen und eine Zeile mit den Freiheitsgraden der Residuen. Bei multivariaten Modellen gibt es ähnliche Zeilen für jede abhängige Variable. Wenn Heteroskedastizität-konsistente Statistiken (nur für Univariate-Modelle verfügbar) ausgewählt wird, wird die Varianz-Kovarianz-Matrix mit einem leistungsfähigen Schätzer berechnet, die Zeile mit Standardfehlern zeigt die stabilen Standardfehler an, und die Signifikanzwerte spiegeln die stabilen Fehler wider. Sie können diese Matrixdatei auch in anderen Prozeduren verwenden, die Matrixdateien einlesen.

## Multivariat: Optionen

In diesem Dialogfeld sind weitere Statistiken verfügbar. Diese werden auf der Grundlage eines Modells mit festen Effekten berechnet.

**Anzeigen.** Mit der Option **Deskriptive Statistiken** lassen Sie beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen berechnen. Die Option **Schätzungen der Effektgröße** liefert einen partiellen Eta-Quadrat-Wert für jeden Effekt und jede Parameterschätzung. Die Eta-Quadrat-Statistik beschreibt den Anteil der Gesamtvariabilität, der einem Faktor zugeschrieben werden kann. Die Option **Beobachtete Trennschärfe** liefert die Testschärfe, wenn die alternative Hypothese auf der Grundlage der beobachteten Werte aufgestellt wurde. Mit **Parameterschätzungen** werden Parameterschätzungen, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Trennschärfe für jeden Test berechnet. Sie können die **SSCP-Matrizen** für die Hypothese und den Fehler, die **SSCP-Matrix der Residuen** sowie den Bartlett-Test auf Sphärizität für die Kovarianzmatrix der Residuen anzeigen lassen.

Mit der Option **Homogenitätstest** wird der Levene-Test auf Homogenität der Varianzen für alle abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjekt Faktoren durchgeführt (nur für Zwischensubjekt Faktoren). Die Homogenitätstests beinhalten auch den Box-*M*-Test auf Homogenität der Kovarianzmatrizen für die abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjekt Faktoren. Die Optionen für Diagramme der Streubreite gegen das mittlere Niveau und Residuendiagramme sind beim Überprüfen von Annahmen über die Daten nützlich. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind. Wählen Sie **Residuendiagramme**, wenn Sie für jede abhängige Variable ein Residuendiagramm (beobachtete über vorhergesagte über standardisierte Werte) erhalten möchten. Diese Diagramme sind beim Überprüfen der Annahme von Gleichheit der Varianzen nützlich. Mit der Option **Test auf fehlende Anpassung** kann überprüft werden, ob die Beziehung zwischen der abhängigen Variablen und der unabhängigen Variablen durch das Modell hinreichend beschrieben werden kann. Die Option **Allgemeine schätzbare Funktion(en)** ermöglicht Ihnen, einen benutzerdefinierten Hypothesentest zu entwickeln, dessen Grundlage die allgemeinen schätzbaren Funktionen sind. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemein schätzbaren Funktionen.

### Anzeige

#### **Deskriptive Statistiken**

Erstellt beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen.

#### **Schätzungen der Effektgröße**

Liefert einen partiellen Eta-Quadrat-Wert für jeden Effekt und jede Parameterschätzung. Die Eta-Quadrat-Statistik beschreibt den Anteil der Gesamtvariabilität, der einem Faktor zugeschrieben werden kann.

#### **Beobachtete Trennschärfe**

Liefert die Testschärfe, wenn die alternative Hypothese auf der Grundlage der beobachteten Werte aufgestellt wurde.

#### **Parameterschätzer**

Berechnet Parameterschätzungen, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Trennschärfe für jeden Test.

#### **SSCP-Matrizen**

Zeigt die Hypothesen und Fehler-SSCP-Matrizen an.

### **SSCP-Matrix für Residuen**

Zeigt die Hypothesen und Fehler-SSCP-Matrizen der Residuen an.

### **Transformationsmatrix**

Zeigt den Bartlett-Test auf Sphärizität für die Kovarianzmatrix der Residuen an.

### **Homogenitätstests**

Führt den Levene-Test auf Homogenität der Varianzen für alle abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjektfaktoren durch (nur für Zwischensubjektfaktoren). Die Homogenitätstests beinhalten auch den Box-M-Test auf Homogenität der Kovarianzmatrizen für die abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjektfaktoren.

### **Diagramm: Streubreite gegen das mittlere Niveau**

Nützlich beim Überprüfen der Annahmen über Daten zum Untersuchen der Annahme von Gleichheit der Varianzen. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind.

### **Residuendiagramm**

Erstellt für jede abhängige Variable ein Residuendiagramm (beobachtete über vorhergesagte über standardisierte Werte) erhalten möchten. Das Diagramm ist beim Überprüfen der Annahme von Gleichheit der Varianzen nützlich.

### **Fehlende Anpassung**

Überprüfen Sie, ob das Modell die Beziehung zwischen der abhängigen Variablen und der unabhängigen Variablen richtig beschreiben kann.

### **Allgemeine schätzbare Funktion(en)**

Hiermit können Sie benutzerdefinierte Hypothesentests entwickeln, deren Grundlage die allgemeine(n) schätzbare(n) Funktion(en) ist bzw. sind. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemein schätzbaren Funktionen.

### **Signifikanzstufe**

Hier können Sie das in den Post-hoc-Tests verwendete Signifikanzniveau und das beim Berechnen von Konfidenzintervallen verwendete Konfidenzniveau ändern. Der hier festgelegte Wert wird auch zum Berechnen der beobachteten Trennschärfe für die Tests verwendet. Wenn Sie ein Signifikanzniveau festlegen, wird das entsprechende Konfidenzniveau im Dialogfeld angezeigt.

## **Zusätzliche Funktionen beim Befehl GLM**

Diese Funktionen gelten für univariate und multivariate Analysen sowie Analysen mit Messwiederholungen. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl DESIGN können Sie verschachtelte Effekte im Design festlegen.
- Mit dem Unterbefehl TEST können Sie Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl CONTRAST können Sie multiple Kontraste angeben.
- Mit dem Unterbefehl MISSING können Sie benutzerdefiniert fehlende Werte aufnehmen.
- Mit dem Unterbefehl CRITERIA können Sie EPS-Kriterien angeben.
- Erstellen Sie eine angepasste **L**-Matrix, eine **M**-Matrix oder eine **K**-Matrix (unter Verwendung der Unterbefehle LMATRIX, MMATRIX oder KMATRIX).
- Mit dem Unterbefehl CONTRAST können Sie bei einfachen und Abweichungskontrasten eine Referenzkategorie zwischenschalten.
- Mit dem Unterbefehl CONTRAST können Sie bei polynomialen Kontrasten Metriken angeben.
- Mit dem Unterbefehl POSTHOC können Sie Fehlerterme für Post-hoc-Vergleiche angeben.
- Mit dem Unterbefehl EMMEANS können Sie geschätzte Randmittel für alle Faktoren oder Interaktionen zwischen den Faktoren in der Faktorenliste berechnen lassen.
- Mit dem Unterbefehl SAVE können Sie Namen für temporäre Variablen angeben.
- Mit dem Unterbefehl OUTFILE können Sie eine Datendatei mit einer Korrelationsmatrix erstellen.

- Mit dem Unterbefehl OUTFILE können Sie eine Matrixdatendatei erstellen, die Statistiken aus der Zwischensubjekt-ANOVA-Tabelle enthält.
  - Mit dem Unterbefehl OUTFILE können Sie die Designmatrix in einer neuen Datendatei speichern.
- Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## GLM - Messwiederholungen

Mithilfe der Prozedur "GLM - Messwiederholungen" können Sie Varianzanalysen durchführen, wenn dieselbe Messung mehrmals für das gleiche Subjekt oder den gleichen Fall vorgenommen wird. Wenn Sie Zwischensubjektfaktoren festlegen, wird die Grundgesamtheit durch diese in Gruppen unterteilt. Mit dieser auf dem allgemeinen linearen Modell basierenden Prozedur können Sie Nullhypothesen über die Effekte der Zwischensubjektfaktoren und die Innersubjektfaktoren testen. Sie können sowohl die Interaktionen zwischen den Faktoren als auch die Auswirkungen einzelner Faktoren untersuchen. Zusätzlich können die Effekte von konstanten Kovariaten und die Interaktionen von Kovariaten mit den Zwischensubjektfaktoren berücksichtigt werden.

In einem Design mit doppelt multivariaten Messwiederholungen stellen die abhängigen Variablen Messungen von mehreren Variablen für die verschiedenen Stufen der Zwischensubjektfaktoren dar. So könnten Sie beispielsweise Puls und Atmung pro Subjekt zu drei verschiedenen Zeiten gemessen haben.

Die Prozedur "GLM - Messwiederholungen" stellt sowohl univariate als auch multivariate Analysen für Daten aus Messwiederholungen zur Verfügung. Es können sowohl ausgewogene als auch unausgewogene Modelle getestet werden. Ein Design ist ausgewogen, wenn jede Zelle im Modell dieselbe Anzahl von Fällen enthält. In einem multivariaten Modell liegen die Quadratsummen aus den Effekten im Modell und die Fehlerquadratsummen in Matrixform und nicht wie bei der univariaten Analyse in skalarer Form vor. Diese Matrizen werden als SSCP-Matrizen (Matrizen der Quadratsummen und Kreuzprodukte) bezeichnet. Mit der Prozedur "GLM - Messwiederholungen" werden nicht nur Hypothesen getestet, sondern zugleich Parameter geschätzt.

Zum Testen von Hypothesen über Zwischensubjektfaktoren stehen häufig verwendete *A-priori*-Kontraste zur Verfügung. Nachdem die Signifikanz mit einem *F*-Gesamttest nachgewiesen wurde, können Sie Post-hoc-Tests verwenden, um Differenzen zwischen bestimmten Mittelwerten berechnen zu lassen. Geschätzte Randmittel dienen als Schätzungen für die vorhergesagten Mittelwerte der Zellen im Modell und mit Profilplots (Interaktionsdiagrammen) dieser Mittelwerte können Sie einige dieser Beziehungen in einfacher Weise visuell darstellen.

Residuen, Einflusswerte, die Cook-Distanz und Hebelwerte können zum Überprüfen von Annahmen als neue Variablen in der Datendatei gespeichert werden. Weiterhin stehen eine SSCP-Matrix für Residuen (eine quadratische Matrix von Quadratsummen und Kreuzprodukten für Residuen), eine Residuenkovarianzmatrix (die SSCP-Matrix für Residuen geteilt durch die Freiheitsgrade der Residuen) und die Residuenkorrelationsmatrix (die standardisierte Form der Residuenkovarianzmatrix) zur Verfügung.

Mit der WLS-Gewichtung können Sie eine Variable angeben, um Beobachtungen für eine WLS-Analyse (Weighted Least Squares - gewichtete kleinste Quadrate) unterschiedlich zu gewichten. Dies kann notwendig sein, um etwaige Unterschiede in der Präzision von Messungen auszugleichen.

**Beispiel.** Zwölf Studenten werden auf der Grundlage der Scores eines Tests in eine Gruppe mit hoher und eine Gruppe mit niedriger Ängstlichkeit eingeteilt. Der ermittelte Wert für die Ängstlichkeit wird "Zwischensubjektfaktor" genannt, da er die Subjekte in Gruppen unterteilt. Die Studenten verfügen über jeweils vier Versuche für eine zu lösende Aufgabe und die Anzahl der Fehler für jeden Versuch wird aufgezeichnet. Die Fehler für jeden Versuch werden in separaten Variablen aufgezeichnet und es wird ein Innersubjektfaktor (Versuch) mit vier Stufen für die vier Versuche definiert. Es stellt sich heraus, dass der Effekt eines Versuchs signifikant ist, während die Interaktion von Versuch und Ängstlichkeit nicht signifikant ist.

**Methoden.** Zum Überprüfen der verschiedenen Hypothesen können Quadratsummen vom Typ I, Typ II, Typ III und Typ IV verwendet werden. Die Voreinstellung sieht den Typ III vor.

**Statistiken.** Post-hoc-Spannweitentests und Mehrfachvergleiche: geringste signifikante Differenz, Bonferroni, Sidak, Scheffé, multiples *F* nach Ryan-Einot-Gabriel-Welsch, multiple Spannweite nach Ryan-

EinotGabriel-Welsch, Student-Newman-Keuls-Test, Tukey-HSD-Test, Tukey-B, Duncan, GT2 nach Hochberg, Gabriel, Waller-Duncan-T-Test, Dunnett (einseitig und zweiseitig), Tamhane-T2, Dunnett-T3, Games-Howell und Dunnett-C. Deskriptive Statistiken: beobachtete Mittelwerte, Standardabweichungen und Anzahl aller abhängigen Variablen in allen Zellen; Levene-Tests auf Homogenität der Varianzen; Box-M-Test und Test auf Sphärizität nach Mauchly.

**Diagramme.** Streubreite gegen mittleres Niveau, Residuendiagramme, Profilplots (Interaktion).

Erläuterungen der Daten für GLM - Messwiederholungen

**Daten.** Die abhängigen Variablen müssen quantitativ sein. Zwischensubjektfaktoren unterteilen die Stichprobe in diskrete Untergruppen, beispielsweise "männlich" und "weiblich". Diese Faktoren sind kategorial und können numerische Werte oder Zeichenfolgewerte besitzen. Innersubjektfaktoren werden im Dialogfeld "Messwiederholungen: Faktor(en) definieren" definiert. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei einer Analyse mit Messwiederholungen müssen diese für jede Stufe der Innersubjektvariablen konstant bleiben.

Die Datendatei muss für jede Gruppe von Messungen am Subjekt ein Set von Variablen enthalten. Jedes Set enthält pro Messwiederholung in der Gruppe eine Variable. Für die Gruppe wird ein Innersubjektfaktor mit einer Anzahl von Stufen, die gleich der Anzahl der Wiederholungen ist, definiert. So könnten beispielsweise an mehreren Tagen Gewichtsmessungen vorgenommen werden. Wenn an fünf Tagen eine Messung derselben Eigenschaft vorgenommen würde, könnte der Innersubjektfaktor als *Tag* mit fünf Stufen definiert werden.

Bei mehreren Innersubjektfaktoren ist die Anzahl der Messungen für jedes Subjekt gleich dem Produkt der Anzahl von Stufen für jeden Faktor. Wenn beispielsweise vier Tage lang jeden Tag drei Messungen vorgenommen werden, dann werden für jedes Subjekt insgesamt zwölf Messungen vorgenommen. Die Innersubjektfaktoren könnten als *Tag(4)* und *Zeit(3)* angegeben werden.

**Annahmen.** Analysen mit Messwiederholungen können auf zwei Arten durchgeführt werden, univariat und multivariat.

Beim univariaten Ansatz (auch als Split-Plot- oder Mixed-Model-Ansatz bekannt) werden die abhängigen Variablen als Reaktionen auf die Stufen der Innersubjektfaktoren betrachtet. Die Messungen an einem Subjekt müssen einer Stichprobe aus einer multivariaten Normalverteilung entstammen und die Varianz-Kovarianz-Matrizen sind für alle durch die Zwischensubjekteffekte gebildeten Zellen gleich. Über die Varianz-Kovarianz-Matrix der abhängigen Variablen werden gewisse Annahmen getroffen. Die Gültigkeit der beim univariaten Ansatz verwendeten *F*-Statistik kann angenommen werden, wenn die Varianz-Kovarianz-Matrix kreisförmig ist (Huynh und Mandeville, 1979).

Um diese Annahme zu testen, kann der Mauchly-Test auf Sphärizität verwendet werden, der einen Test auf Sphärizität für die Varianz-Kovarianz-Matrix einer orthonormalisierten, transformierten abhängigen Variablen durchführt. Der Mauchly-Test wird bei Analysen mit Messwiederholungen automatisch angezeigt. Bei kleinen Stichprobengrößen ist dieser Test nicht sehr leistungsfähig. Bei großen Stichprobengrößen kann der Test sogar dann signifikant sein, wenn der Einfluss der Abweichung auf die Ergebnisse klein ist. Wenn die Signifikanz des Tests groß ist, kann die Hypothese der Sphärizität angenommen werden. Wenn die Signifikanz jedoch klein ist und die Annahme der Sphärizität verletzt zu sein scheint, kann eine Anpassung der Zähler- und Nennerfreiheitsgrade vorgenommen werden, um die univariate *F*-Statistik zu bestätigen. In der Prozedur "GLM - Messwiederholungen" sind drei Schätzungen für diese auch **Epsilon** genannte Anpassung vorhanden. Sowohl die Zähler- als auch die Nennerfreiheitsgrade müssen mit Epsilon multipliziert werden und die Signifikanz des *F*-Quotienten muss unter Berücksichtigung der neuen Freiheitsgrade bewertet werden.

Beim multivariaten Ansatz werden die Messungen für ein Subjekt als Stichprobe aus einer multivariaten Normalverteilung angesehen und die Varianz-Kovarianz-Matrizen sind für alle durch die Zwischensubjektfaktoren gebildeten Zellen gleich. Für den Test, ob die Varianz-Kovarianz-Matrizen für die Zellen gleich sind, kann der Box-M-Test verwendet werden.

**Verwandte Prozeduren.** Verwenden Sie die Prozedur "Explorative Datenanalyse" zum Untersuchen der Daten, ehe Sie eine Varianzanalyse durchführen. Wenn es *nicht* wiederholte Messungen für jedes Subjekt gibt, verwenden Sie die Prozeduren "GLM - Univariat" oder "GLM - Multivariat". Wenn für jedes Subjekt nur zwei Messungen vorhanden sind (zum Beispiel vor und nach einem Test) und es keine Zwischensubjektfaktoren gibt, können Sie die Prozedur "t-Test bei Stichproben mit paarigen Werten" verwenden.

Ermittlung von "GLM - Messwiederholungen"

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Allgemeines lineares Modell > Messwiederholung...**

2. Geben Sie den Namen eines Innersubjektfaktors und die Anzahl seiner Stufen ein.

3. Klicken Sie auf **Add**.

4. Wiederholen Sie diesen Vorgang für jeden Innersubjektfaktor.

So definieren Sie Messwertfaktoren für ein Design mit doppelt multivariaten Messwiederholungen:

5. Geben Sie die Messwertbezeichnung ein.

6. Klicken Sie auf **Add**.

Nachdem Sie alle Faktoren und Messwerte definiert haben:

7. Klicken Sie auf **Definieren**.

8. Wählen Sie eine abhängige Variable aus, die jeder Kombination von Innersubjektfaktoren (und wahlweise Messwerten) in der Liste entspricht.

Sie können die Positionen dieser Variablen ändern, indem Sie auf die nach oben bzw. die nach unten weisende Pfeilschaltfläche klicken.

Sie können Änderungen an den Innersubjektfaktoren vornehmen, indem Sie das Dialogfeld "Messwiederholungen: Faktor(en) definieren" erneut öffnen. Hierfür muss das Hauptdialogfeld nicht geschlossen werden. Wahlweise können Sie Zwischensubjektfaktoren und Kovariaten angeben.

## Messwiederholungen: Faktor(en) definieren

Die Prozedur "GLM - Messwiederholungen" analysiert Gruppen von miteinander in Beziehung stehenden, abhängigen Variablen, die verschiedene Messungen desselben Attributs darstellen. In diesem Dialogfeld können Sie einen oder mehrere Innersubjektfaktoren definieren, die in der Prozedur "GLM - Messwiederholungen" verwendet werden. Beachten Sie, dass die Reihenfolge wichtig ist, in der Sie die Innersubjektfaktoren festlegen. Jeder Faktor bildet eine Stufe innerhalb des vorhergehenden Faktors.

Damit Sie die Prozedur "GLM - Messwiederholungen" verwenden können, müssen Sie die Daten korrekt anlegen. In diesem Dialogfeld müssen Sie Innersubjektfaktoren definieren. Beachten Sie, dass diese Faktoren keine Variablen in den Daten sind, sondern Faktoren, die Sie hier definieren.

**Beispiel.** In einer Diätstudie wird das Gewicht von mehreren Personen fünf Wochen lang einmal wöchentlich gemessen. In der Datendatei wird jede Person durch ein Subjekt (also einen Fall) dargestellt. Die Gewichte für die einzelnen Wochen werden in den Variablen *Gewicht1*, *Gewicht2* usw. aufgezeichnet. Das Geschlecht der einzelnen Personen wird in einer weiteren Variablen vermerkt. Die für jedes Subjekt wiederholt gemessenen Gewichte können durch Definieren eines Innersubjektfaktors gruppiert werden. Dieser Faktor kann als *Woche* bezeichnet werden, dem fünf Stufen zugeordnet werden. Im Hauptdialogfeld werden die Variablen *Gewicht1* bis *Gewicht5* zum Zuordnen der fünf Stufen zu *Woche* verwendet. Die Variable in der Datendatei, die Männer und Frauen gruppiert (*Geschl*), kann als Zwischensubjektfaktor angegeben werden, um die Unterschiede zwischen den Männern und Frauen zu untersuchen.

**Messungen.** Wenn bei den Subjekten jeweils mehr als ein Messwert gemessen wurde, müssen Sie die Messwerte definieren. So könnten zum Beispiel bei jedem Subjekt eine Woche lang täglich Puls und Atmung gemessen werden. Diese Messwerte sind nicht als Variablen in der Datendatei vorhanden, sondern werden hier definiert. Modelle mit mehr als einem Messwert werden auch als Modell mit doppelt multivariaten Messwiederholungen bezeichnet.

## Messwiederholungen: Modell

**Modell angeben.** Ein gesättigtes Modell enthält alle Haupteffekte der Faktoren, alle Kovariatenhaupteffekte und alle Interaktionen zwischen Faktoren. Es enthält keine Kovariateninteraktionen. Wählen Sie **Anpassen** aus, um nur ein Subset von Interaktionen oder Interaktionen zwischen Faktoren und Kovariaten festzulegen. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

**Zwischen den Subjekten.** Die Zwischensubjektfaktoren und Kovariaten werden aufgelistet.

**Modell.** Das Modell ist von der Art Ihrer Daten abhängig. Nachdem Sie **Anpassen** ausgewählt haben, können Sie die für die Analyse relevanten Innersubjekteffekte und -interaktionen und die Zwischensubjekteffekte und -interaktionen auswählen.

**Quadratsumme.** Die Methode zum Berechnen der Quadratsummen für das Zwischensubjektmodell. Für ausgewogene und unausgewogene Zwischensubjektmodelle ohne fehlende Zellen wird meistens die Methode mit Quadratsummen vom Typ III angewendet.

## Erstellen von Termen und benutzerdefinierten Termen

### Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie nicht verschachtelte Terme eines bestimmten Typs (beispielsweise Haupteffekte) für alle Kombinationen eines ausgewählten Sets von Faktoren und Kovariaten einschließen wollen.

### Benutzerdefinierte Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie verschachtelte Terme einschließen oder einen Term explizit Variable für Variable erstellen wollen. Die Erstellung eines verschachtelten Terms besteht aus folgenden Schritten:

## Quadratsumme

Für das Modell können Sie einen Typ von Quadratsumme auswählen. Typ III wird am häufigsten verwendet und ist die Standardeinstellung.

**Typ I.** Diese Methode ist auch als die Methode der hierarchischen Zerlegung der Quadratsummen bekannt. Jeder Term wird nur für den Vorläuferterm im Modell korrigiert. Quadratsummen vom Typ I werden gewöhnlich in den folgenden Situationen verwendet:

- Ein ausgewogenes ANOVA-Modell, in dem alle Haupteffekte vor den Interaktionseffekten 1. Ordnung festgelegt werden, alle Interaktionseffekte 1. Ordnung wiederum vor den Interaktionseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme niedrigerer Ordnung vor den Termen höherer Ordnung festgelegt werden.
- Ein rein verschachteltes Modell, in welchem der zuerst bestimmte Effekt in dem als zweiten bestimmten Effekt verschachtelt ist, der zweite Effekt wiederum im dritten und so weiter. (Diese Form der Verschachtelung kann nur durch Verwendung der Befehlssprache erreicht werden.)

**Typ II.** Bei dieser Methode wird die Quadratsumme eines Effekts im Modell angepasst an alle anderen "zutreffenden" Effekte berechnet. Ein zutreffender Effekt ist ein Effekt, der mit allen Effekten in Beziehung steht, die den untersuchten Effekt nicht enthalten. Die Methode mit Quadratsummen vom Typ II wird gewöhnlich in den folgenden Fällen verwendet:

- Bei ausgewogenen ANOVA-Modellen.
- Bei Modellen, die nur Haupteffekte von Faktoren enthalten.
- Jedes Regressionsmodell.
- Bei rein verschachtelten Designs. (Diese Form der Verschachtelung kann durch Verwendung der Befehlssprache erreicht werden.)

**Typ III.** Der Standardparameter. Bei dieser Methode werden die Quadratsummen eines Effekts im Design als Quadratsummen orthogonal zu allen Effekten (sofern vorhanden), die den Effekt enthalten, und mit Bereinigung um alle anderen Effekte berechnet, die den Effekt nicht enthalten. Der große Vorteil der Quadratsummen vom Typ III ist, dass sie invariant bezüglich der Zellenhäufigkeiten sind, solange die allgemeine Form der Schätzbarkeit konstant bleibt. Daher wird dieser Typ von Quadratsumme oft für unausgewogene Modelle ohne fehlende Zellen als geeignet angesehen. In einem faktoriellen Design ohne fehlende Zellen ist diese Methode äquivalent zu der Methode der gewichteten Mittelwertquadrate nach Yates. Die Methode mit Quadratsummen vom Typ III wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.

- Alle ausgewogenen oder unausgewogenen Modelle ohne leere Zellen.

**Typ IV.** Diese Methode ist dann geeignet, wenn es keine fehlenden Zellen gibt. Für alle Effekte  $F$  im Design: Wenn  $F$  in keinem anderen Effekt enthalten ist, dann gilt: Typ IV = Typ III = Typ II. Wenn  $F$  in anderen Effekten enthalten ist, werden bei Typ IV die Kontraste zwischen den Parametern in  $F$  gleichmäßig auf alle Effekte höherer Ordnung verteilt. Die Methode mit Quadratsummen vom Typ IV wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgewogenen oder unausgewogenen Modelle mit leeren Zellen.

## Messwiederholungen: Kontraste

Kontraste werden verwendet, um auf Unterschiede zwischen den Stufen eines Zwischensubjektfaktors zu testen. Sie können für jeden Zwischensubjektfaktor im Modell einen Kontrast festlegen. Kontraste stellen lineare Kombinationen der Parameter dar.

Das Testen von Hypothesen basiert auf der Nullhypothese  $\mathbf{LBM} = 0$ , wobei  $\mathbf{L}$  die Kontrastkoeffizientenmatrix,  $\mathbf{B}$  der Parametervektor und  $\mathbf{M}$  die gemittelte Matrix ist, die der Mittelwerttransformation der abhängigen Variablen entspricht. Sie können diese Transformationsmatrix anzeigen lassen, indem Sie im Dialogfeld "Messwiederholungen: Optionen" die Option **Transformationsmatrix** auswählen. Wenn es zum Beispiel vier abhängige Variablen und einen Innersubjektfaktor mit vier Stufen gibt und polynomiale Kontraste (die Voreinstellung) für Innersubjektfaktoren verwendet werden, dann entspricht die  $\mathbf{M}$ -Matrix  $(0,5 \ 0,5 \ 0,5 \ 0,5)'$ . Wenn Sie einen Kontrast festlegen, wird eine  $\mathbf{L}$ -Matrix erstellt, bei der die Spalten für den betreffenden Zwischensubjektfaktor mit dem Kontrast übereinstimmen. Die verbleibenden Spalten werden so angepasst, dass die  $\mathbf{L}$ -Matrix schätzbar ist.

Als Kontraste sind "Abweichung", "Einfach", "Differenz", "Helmert", "Wiederholt" und "Polynomial" verfügbar. Bei Abweichungskontrasten und einfachen Kontrasten können Sie wählen, ob die letzte oder die erste Kategorie als Referenzkategorie dient.

Für Innersubjektfaktoren müssen Sie einen anderen Kontrast als **Keine** auswählen.

### Kontrasttypen

**Abweichung.** Vergleicht den Mittelwert jeder Faktorstufe (außer bei Referenzkategorien) mit dem Mittelwert aller Faktorstufen (Gesamtmittelwert). Die Stufen des Faktors können in beliebiger Ordnung vorliegen.

**Einfach.** Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt. Sie können die erste oder die letzte Kategorie als Referenz auswählen.

**Differenz.** Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. (Dies wird gelegentlich auch als umgekehrter Helmert-Kontrast bezeichnet).

**Helmert.** Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.

**Wiederholt.** Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.

**Polynomial.** Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

## GLM - Messwiederholung: Profilplots

Profilplots (Interaktionsdiagramme) sind hilfreich zum Vergleichen von Randmitteln im Modell. Ein Profilplot ist ein Liniendiagramm, in dem jeder Punkt das geschätzte Randmittel einer abhängigen Variablen (angepasst an die Kovariaten) bei einer Stufe eines Faktors angibt. Die Stufen eines zweiten Faktors können zum Erzeugen getrennter Linien verwendet werden. Jede Stufe in einem dritten Faktor kann verwendet werden, um ein separates Diagramm zu erzeugen. Für Plots stehen alle Faktoren zur Verfügung.

Profilplots werden für jede einzelne abhängige Variable erstellt. In Profilplots können sowohl Zwischen-subjektfaktoren als auch Innersubjektfaktoren verwendet werden.

Ein Profilplot für einen Faktor zeigt, ob die geschätzten Randmittel mit den Faktorstufen steigen oder fallen. Bei zwei oder mehr Faktoren deuten parallele Linien an, dass es keine Interaktion zwischen den Faktoren gibt. Das heißt, dass Sie die Faktorstufen eines einzelnen Faktors untersuchen können. Nicht parallele Linien deuten auf eine Interaktion hin.

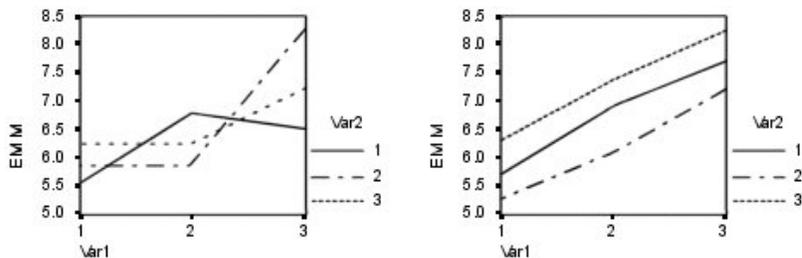


Abbildung 2. Nicht paralleles Diagramm (links) und paralleles Diagramm (rechts)

Nachdem ein Diagramm durch Auswahl von Faktoren für die horizontale Achse (und wahlweise von Faktoren für getrennte Linien und getrennte Diagramme) festgelegt wurde, muss das Diagramm der Liste "Diagramme" hinzugefügt werden.

## GLM - Messwiederholung: Post-hoc-Mehrfachvergleiche

**Tests für Post-hoc-Mehrfachvergleiche.** Sobald Sie festgestellt haben, dass es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-hoc-Spannweitentests und paarweisen multiplen Vergleichen untersuchen, welche Mittelwerte sich unterscheiden. Die Vergleiche werden auf der Basis von nicht korrigierten Werten vorgenommen. Diese Tests sind nicht verfügbar, wenn es keine Zwischensubjektfaktoren gibt, und die Post-hoc-Mehrfachvergleiche werden für den Durchschnitt aller Stufen der Innersubjektfaktoren durchgeführt.

Häufig verwendete Mehrfachvergleiche sind der Bonferroni-Test und der Tukey-HSD-Test. Der **Bonferroni-Test** auf der Grundlage der studentisierten  $T$ -Statistik korrigiert das beobachtete Signifikanzniveau unter Berücksichtigung der Tatsache, dass multiple Vergleiche vorgenommen werden. Der **Sidak-t-Test** korrigiert ebenfalls das Signifikanzniveau und liefert engere Grenzen als der Bonferroni-Test. Der **Tukey-HSD-Test** verwendet die studentisierte Spannweitenstatistik, um alle paarweisen Vergleiche zwischen den Gruppen vorzunehmen, und setzt die experimentelle Fehlerrate auf die Fehlerrate der Ermittlung aller paarweisen Vergleiche. Beim Testen einer großen Anzahl von Mittelwertpaaren ist der Tukey-HSD-Test leistungsfähiger als der Bonferroni-Test. Bei einer kleinen Anzahl von Paaren ist der Bonferroni-Test leistungsfähiger.

**GT2 nach Hochberg** ähnelt dem Tukey-HSD-Test, es wird jedoch der studentisierte Maximalmodulus verwendet. Meistens ist der Tukey-HSD-Test leistungsfähiger. Der **paarweise Vergleichstest nach Gabriel** verwendet ebenfalls den studentisierten Maximalmodulus und zeigt meistens eine größere Schärfe als das GT2 nach Hochberg, wenn die Zellengrößen ungleich sind. Der Gabriel-Test kann ungenau werden, wenn die Zellengrößen stark variieren.

Mit dem **paarweisen t-Test für mehrere Vergleiche nach Dunnett** wird ein Set von Verarbeitungen mit einem einzelnen Kontrollmittelwert verglichen. Als Kontrollkategorie ist die letzte Kategorie voreingestellt. Sie können aber auch die erste Kategorie einstellen. Außerdem können Sie einen einseitigen oder zweiseitigen Test wählen. Verwenden Sie einen zweiseitigen Test, um zu überprüfen, ob sich der Mittelwert bei jeder Stufe (außer der Kontrollkategorie) des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet. Um zu überprüfen, ob sich der Mittelwert bei jeder Stufe des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet, wählen Sie **< Control**. Wählen Sie **>Steuerung** aus, um zu testen, ob der Mittelwert auf einer beliebigen Ebene des Faktors größer ist als die der Kontrollkategorie.

Ryan, Einot, Gabriel und Welsch (R-E-G-W) entwickelten zwei multiple Step-down-Spannweitentests. Multiple Step-down-Prozeduren überprüfen zuerst, ob alle Mittelwerte gleich sind. Wenn nicht alle Mittelwerte gleich sind, werden Subsets der Mittelwerte auf Gleichheit getestet. Das **F nach R-E-G-W** basiert auf einem  $F$ -Test und **Q nach R-E-G-W** basiert auf der studentisierten Spannweite. Diese Tests sind

leistungsfähiger als der multiple Spannweitentest nach Duncan und der Student-Newman-Keuls-Test (ebenfalls multiple Step-down-Prozeduren), aber sie sind bei ungleichen Zellengrößen nicht empfehlenswert.

Bei ungleichen Varianzen verwenden Sie **Tamhane-T2** (konservativer paarweiser Vergleichstest auf der Grundlage eines *T*-Tests), **Dunnnett-T3** (paarweiser Vergleichstest auf der Grundlage des studentisierten Maximalmodulus), den **paarweisen Vergleichstest** nach **Games-Howell** (manchmal ungenau) oder das **Dunnnett-C** (paarweiser Vergleichstest auf der Grundlage der studentisierten Spannweite).

Der **multiple Spannweitentest nach Duncan**, Student-Newman-Keuls (**S-N-K**) und **Tukey-B** sind Spannweitentests, mit denen Gruppenmittelwerte geordnet und ein Wertebereich berechnet wird. Diese Tests werden nicht so häufig verwendet wie die vorher beschriebenen Tests.

Der **Waller-Duncan-t-Test** verwendet die Bayes-Methode. Dieser Spannweitentest verwendet den harmonischen Mittelwert der Stichprobengröße, wenn die Stichprobengrößen ungleich sind.

Das Signifikanzniveau des **Scheffé**-Tests ist so festgelegt, dass alle möglichen linearen Kombinationen von Gruppenmittelwerten getestet werden können und nicht nur paarweise Vergleiche verfügbar sind, wie bei dieser Funktion der Fall. Das führt dazu, dass der Scheffé-Test oftmals konservativer als andere Tests ist, also für eine Signifikanz eine größere Differenz der Mittelwerte erforderlich ist.

Der paarweise multiple Vergleichstest auf geringste signifikante Differenz (**LSD**) ist äquivalent zu multiplen individuellen *T*-Tests zwischen allen Gruppenpaaren. Der Nachteil bei diesem Test ist, dass kein Versuch unternommen wird, das beobachtete Signifikanzniveau im Hinblick auf multiple Vergleiche zu korrigieren.

**Angezeigte Tests.** Es werden paarweise Vergleiche für LSD, Sidak, Bonferroni, Games-Howell, T2 und T3 nach Tamhane, Dunnnett-C und Dunnnett-T3 ausgegeben. Homogene Subsets für Spannweitentests werden ausgegeben für S-N-K, Tukey-B, Duncan, *F* nach R-E-G-W, *Q* nach R-E-G-W und Waller. Der Tukey-HSD-Test, das GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind multiple Vergleiche, zugleich aber auch Spannweitentests.

## GLM: Geschätzte Randmittel

Wählen Sie die Faktoren und Interaktionen aus, für die Sie Schätzungen für die Randmittel der Grundgesamtheit in den Zellen wünschen. Diese Mittel werden gegebenenfalls an die Kovariaten angepasst.

### Haupteffekte vergleichen

Gibt nicht korrigierte paarweise Vergleiche zwischen den geschätzten Randmitteln für alle Haupteffekte im Modell aus, sowohl für Zwischensubjektfaktoren als auch für Innersubjektfaktoren. Diese Option ist nur verfügbar, falls in der Liste "Mittelwerte anzeigen für" Haupteffekte ausgewählt sind.

### Einfache Haupteffekte vergleichen

Diese Einstellung wird immer dann aktiviert, wenn die Zielliste mindestens einen Produkt- oder Interaktionseffekt enthält (zum Beispiel  $A \times B$ ,  $A \times B \times C$ ). Die Einstellung unterstützt die Spezifikation von Vergleichen bei einfachen Haupteffekten, bei denen es sich um Haupteffekte handelt, die innerhalb der Stufen anderer Faktoren verschachtelt sind.

### Anpassung des Konfidenzintervalls

Wählen Sie für das Konfidenzintervall und die Signifikanz entweder die geringste signifikante Differenz (LSD - Least Significant Difference), Bonferroni oder die Anpassung nach Sidak. Diese Option ist nur verfügbar, wenn **Haupteffekte vergleichen** und/oder **Einfache Haupteffekte vergleichen** ausgewählt ist.

## Angeben geschätzter Randmittel

1. Wählen Sie aus den Menüs eine der unter **> Analysieren > Allgemeines lineares Modell** verfügbaren Prozeduren aus.
2. Klicken Sie im Hauptdialogfeld auf **Geschätzte Randmittel**.

## GLM - Messwiederholungen: Speichern

Vom Modell vorhergesagte Werte, Residuen und verwandte Maße können als neue Variablen im Dateneditor gespeichert werden. Viele dieser Variablen können zum Untersuchen von Annahmen über die Daten verwendet werden. Um die Werte zur Verwendung in einer anderen IBM SPSS Statistics-Sitzung zu speichern, müssen Sie die aktuelle Datendatei speichern.

**Vorhergesagte Werte.** Dies sind die Werte, welche das Modell für die einzelnen Fälle vorhersagt.

- *Nicht standardisiert.* Der Wert, den das Modell für die abhängige Variable vorhersagt.
- *Standardfehler.* Ein Schätzwert der Standardabweichung des Durchschnittswertes der abhängigen Variablen für die Fälle, die dieselben Werte für die unabhängigen Variablen haben.

**Diagnose.** Dies sind Maße zum Auffinden von Fällen mit ungewöhnlichen Wertekombinationen bei den unabhängigen Variablen und von Fällen, die einen großen Einfluss auf das Modell haben könnten. Zur Verfügung stehen die Cook-Distanz und nicht zentrierte Hebelwerte.

- *Cook-Distanz.* Ein Maß dafür, wie sehr sich die Residuen aller Fälle ändern würden, wenn ein bestimmter Fall aus der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, dass der Ausschluss eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substantiell verändert.
- *Hebelwerte.* Nicht zentrierte Hebelwerte. Der relative Einfluss einer jeden Beobachtung auf die Anpassungsgüte eines Modells.

**Residuen.** Ein nicht standardisiertes Residuum ist der tatsächliche Wert der abhängigen Variablen minus des vom Modell geschätzten Werts. Ebenfalls verfügbar sind standardisierte, studentisierte und ausgeschlossene Residuen.

- *Nicht standardisiert.* Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- *Standardisiert.* Das Residuum dividiert durch eine Schätzung der Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- *Studentisiert.* Das Residuum dividiert durch eine Schätzung seiner Standardabweichung, die von Fall zu Fall variiert, abhängig von der Entfernung der Werte jedes Falls auf den unabhängigen Variablen vom Mittelwert der unabhängigen Variablen. Wird manchmal als intern studentisiertes Residuum bezeichnet.
- *Gelöscht.* Das Residuum für einen Fall, wenn dieser Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen wird. Dies ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.

**Koeffizientenstatistik.** Hiermit können Sie eine Varianz-Kovarianz-Matrix der Parameterschätzungen in ein Dataset oder eine Datendatei speichern. Für jede abhängige Variable gibt es weiterhin eine Zeile mit Parameterschätzungen, eine Zeile mit Signifikanzwerten für die *T*-Statistik der betreffenden Parameterschätzungen und eine Zeile mit den Freiheitsgraden der Residuen. Bei multivariaten Modellen gibt es ähnliche Zeilen für jede abhängige Variable. Sie können diese Matrixdaten auch in anderen Prozeduren verwenden, die Matrixdateien einlesen. Datasets sind für die anschließende Verwendung in der gleichen Sitzung verfügbar, werden jedoch nicht als Dateien gespeichert, sofern Sie diese nicht ausdrücklich vor dem Beenden der Sitzung speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen. .

## GLM - Messwiederholungen: Optionen

In diesem Dialogfeld sind weitere Statistiken verfügbar. Diese werden auf der Grundlage eines Modells mit festen Effekten berechnet.

### Anzeige

#### Deskriptive Statistiken

Erstellt beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen.

### **Schätzungen der Effektgröße**

Liefert einen partiellen Eta-Quadrat-Wert für jeden Effekt und jede Parameterschätzung. Die Eta-Quadrat-Statistik beschreibt den Anteil der Gesamtvariabilität, der einem Faktor zugeschrieben werden kann.

### **Beobachtete Trennschärfe**

Liefert die Testschärfe, wenn die alternative Hypothese auf der Grundlage der beobachteten Werte aufgestellt wurde.

### **Parameterschätzungen**

Berechnet Parameterschätzungen, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Trennschärfe für jeden Test.

### **SSCP-Matrizen**

Zeigt die Hypothesen und Fehler-SSCP-Matrizen an.

### **SSCP-Matrix der Residuen**

Zeigt die SSCP-Matrix der Residuen an.

### **Transformationsmatrix**

Zeigt den Bartlett-Test auf Sphärizität für die Kovarianzmatrix der Residuen an.

### **Homogenitätstests**

Führt den Levene-Test auf Homogenität der Varianzen für alle abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjektfaktoren durch (nur für Zwischensubjektfaktoren). Die Homogenitätstests beinhalten auch den Box-*M*-Test auf Homogenität der Kovarianzmatrizen für die abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjektfaktoren.

### **Streubreite vs. mittleres Niveau**

Nützlich für die Überprüfung von Annahmen über die Daten. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind.

### **Residuendiagramm**

Erstellt für jede abhängige Variable ein Residuendiagramm (beobachtete über vorhergesagte über standardisierte Werte) erhalten möchten. Diese Diagramme sind beim Überprüfen der Annahme von Gleichheit der Varianzen nützlich. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind.

### **Fehlende Anpassung**

Überprüft, ob das Modell die Beziehung zwischen der abhängigen Variablen und der unabhängigen Variablen richtig beschreiben kann.

### **Allgemeine schätzbare Funktion(en)**

Ermöglicht es Ihnen, einen benutzerdefinierten Hypothesentest zu entwickeln, dessen Grundlage die allgemein schätzbaren Funktionen sind. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemeinen schätzbaren Funktion(en).

### **Signifikanzniveau**

Hier können Sie das in den Post-hoc-Tests verwendete Signifikanzniveau und das beim Berechnen von Konfidenzintervallen verwendete Konfidenzniveau ändern. Der hier festgelegte Wert wird auch zum Berechnen der beobachteten Trennschärfe für die Tests verwendet. Wenn Sie ein Signifikanzniveau festlegen, wird das entsprechende Konfidenzniveau im Dialogfeld angezeigt.

## **Zusätzliche Funktionen beim Befehl GLM**

Diese Funktionen gelten für univariate und multivariate Analysen sowie Analysen mit Messwiederholungen. Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl DESIGN können Sie verschachtelte Effekte im Design festlegen.
- Mit dem Unterbefehl TEST können Sie Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl CONTRAST können Sie multiple Kontraste angeben.
- Mit dem Unterbefehl MISSING können Sie benutzerdefiniert fehlende Werte aufnehmen.

- Mit dem Unterbefehl CRITERIA können Sie EPS-Kriterien angeben.
- Erstellen Sie eine angepasste **L**-Matrix, eine **M**-Matrix oder eine **K**-Matrix (unter Verwendung der Unterbefehle LMATRIX, MMATRIX und KMATRIX).
- Mit dem Unterbefehl CONTRAST können Sie bei einfachen und Abweichungskontrasten eine Referenzkategorie zwischenschalten.
- Mit dem Unterbefehl CONTRAST können Sie bei polynomialen Kontrasten Metriken angeben.
- Mit dem Unterbefehl POSTHOC können Sie Fehlerterme für Post-hoc-Vergleiche angeben.
- Mit dem Unterbefehl EMMEANS können Sie geschätzte Randmittel für alle Faktoren oder Interaktionen zwischen den Faktoren in der Faktorenliste berechnen lassen.
- Mit dem Unterbefehl SAVE können Sie Namen für temporäre Variablen angeben.
- Mit dem Unterbefehl OUTFILE können Sie eine Datendatei mit einer Korrelationsmatrix erstellen.
- Mit dem Unterbefehl OUTFILE können Sie eine Matrixdatendatei erstellen, die Statistiken aus der Zwischensubjekt-ANOVA-Tabelle enthält.
- Mit dem Unterbefehl OUTFILE können Sie die Designmatrix in einer neuen Datendatei speichern.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Analyse der Varianzkomponenten

---

Die Prozedur "Varianzkomponenten" schätzt bei Modellen mit gemischten Effekten den Beitrag jedes Zufallseffekts zur Varianz der abhängigen Variablen. Diese Prozedur ist besonders wichtig für Analysen von gemischten Modellen, wie zum Beispiel bei univariaten Messwiederholungen und Designs mit Zufallsblöcken. Durch die Berechnung von Varianzkomponenten können Sie ermitteln, wo Sie die Varianz möglicherweise noch reduzieren können.

Vier verschiedene Methoden sind verfügbar, um die Varianzkomponenten zu schätzen: quadratischer unverzerrter Schätzer mit minimaler Norm (MINQUE), Varianzanalyse (ANOVA), Maximum Likelihood (ML) und eingeschränkte Maximum Likelihood (REML). Für die Methoden können Sie eine Reihe von Festlegungen treffen.

Die Standardausgabe für alle Methoden beinhaltet Schätzungen der Varianzkomponenten. Wenn Sie die ML-Methode oder die REML-Methode verwenden, wird außerdem eine Tabelle der asymptotischen Kovarianzmatrix angezeigt. Weiterhin stehen eine ANOVA-Tabelle und erwartete Mittel der Quadrate für die ANOVA-Methode und ein Iterationsverlauf für die ML- und REML-Methoden zur Verfügung. Die Prozedur "Varianzkomponenten" ist vollständig kompatibel zur Prozedur "GLM - Univariat".

Mit der WLS-Gewichtung können Sie eine Variable festlegen, mit der bei gewichteten Analysen Beobachtungen verschieden gewichtet werden, beispielsweise um Schwankungen in der Messgenauigkeiten zu kompensieren.

**Beispiel.** In einem landwirtschaftlichen Betrieb wird die Gewichtszunahme von Schweinen in sechs verschiedenen Würfen nach einem Monat gemessen. Die Variable "Wurf" ist ein Zufallsfaktor mit sechs Stufen. (Die sechs untersuchten Würfe sind eine Zufallsstichprobe aus einer großen Grundgesamtheit von Würfen.) Es stellt sich heraus, dass die Varianz der Gewichtszunahme viel stärker den Unterschieden bei den Würfen als den Unterschieden bei den Ferkeln innerhalb eines Wurfs zugeschrieben werden kann.

Erläuterungen der Daten für Varianzkomponenten

**Daten.** Die abhängige Variable ist quantitativ. Faktoren sind kategorial. Sie können numerische Werte oder Zeichenfolgewerte von bis zu acht Byte Länge annehmen. Mindestens einer der Faktoren muss zufällig sein. Das heißt, die Stufen des Faktors müssen eine Zufallsstichprobe aus den möglichen Stufen sein. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen.

**Annahmen.** Alle Methoden gehen davon aus, dass die Modellparameter eines Zufallseffekts Mittelwerte gleich null und endliche konstante Varianzen aufweisen und nicht miteinander korrelieren. Auch die Modellparameter von verschiedenen Zufallseffekten korrelieren nicht miteinander.

Der Residuenterm besitzt ebenfalls einen Mittelwert gleich null und eine endliche konstante Varianz. Er korreliert nicht mit den Modellparametern aller Zufallseffekte. Es wird davon ausgegangen, dass Residuenterm aus verschiedenen Beobachtungen nicht korrelieren.

Ausgehend von diesen Annahmen sind Beobachtungen aus denselben Stufen eines Zufallsfaktors korreliert. Diese Tatsache unterscheidet ein Varianzkomponentenmodell von einem allgemeinen linearen Modell.

Bei ANOVA und MINQUE wird keine Normalverteilung vorausgesetzt. Beide Verfahren sind robust gegenüber mäßigen Abweichungen von der Normalverteilung.

Für ML und REML müssen die Modellparameter und der Residuenterm normalverteilt sein.

**Verwandte Prozeduren.** Verwenden Sie die Prozedur "Explorative Datenanalyse" zum Untersuchen der Daten, ehe Sie eine Varianzkomponentenanalyse durchführen. Verwenden Sie zum Testen von Hypothesen die Prozeduren "GLM - Univariat", "GLM - Multivariat" und "GLM - Messwiederholungen".

Berechnen von Varianzkomponentenanalysen

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Allgemeines lineares Modell > Varianzkomponenten...**

2. Wählen Sie eine abhängige Variable aus.

3. Wählen Sie in Abhängigkeit von den Daten Variablen als feste Faktoren, Zufallsfaktoren und Kovariaten aus. Als WLS-Gewichtung können Sie eine Gewichtungsvariable angeben.

## Varianzkomponenten: Modell

**Modell angeben.** Ein gesättigtes Modell enthält alle Haupteffekte der Faktoren, alle Kovariatenhaupteffekte und alle Interaktionen zwischen Faktoren. Es enthält keine Kovariateninteraktionen. Wählen Sie **Anpassen** aus, um nur ein Subset von Interaktionen oder Interaktionen zwischen Faktoren und Kovariaten festzulegen. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

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

**Modell.** Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von **Anpassen** können Sie die Haupteffekte und Interaktionen auswählen, die für Ihre Analyse von Interesse sind. Das Modell muss einen Zufallsfaktor enthalten.

Für die ausgewählten Faktoren und Kovariaten:

### Interaktion

Hiermit wird der Interaktionsterm mit der höchsten Ordnung von allen ausgewählten Variablen erstellt. Dies ist der Standardwert.

### Haupteffekte

Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

### Alle 2-Wege

Hiermit werden alle möglichen Zweibegeinteraktionen der ausgewählten Variablen erstellt.

### Alle 3-Wege

Hiermit werden alle möglichen Dreibegeinteraktionen der ausgewählten Variablen erzeugt.

### Alle 4-Wege

Erstellt alle möglichen Vierbegeinteraktionen der ausgewählten Variablen.

### Alle 5-Wege

Hiermit werden alle möglichen Fünfbegeinteraktionen der ausgewählten Variablen erzeugt.

**Konstanter Term in Modell einschließen.** Der konstante Term wird gewöhnlich in das Modell aufgenommen. Falls Sie sicher sind, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

## Erstellen von Termen und benutzerdefinierten Termen

### Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie nicht verschachtelte Terme eines bestimmten Typs (beispielsweise Haupteffekte) für alle Kombinationen eines ausgewählten Sets von Faktoren und Kovariaten einschließen wollen.

### Benutzerdefinierte Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie verschachtelte Terme einschließen oder einen Term explizit Variable für Variable erstellen wollen. Die Erstellung eines verschachtelten Terms besteht aus folgenden Schritten:

## Varianzkomponenten: Optionen

**Methode.** Sie können eine von vier Methoden zum Schätzen der Varianzkomponenten verwenden.

- Mit **MINQUE** (quadratischer unverzerrter Schätzer mit minimaler Norm) werden Schätzungen berechnet, die invariant in Bezug auf die festen Effekte sind. Wenn die Daten normalverteilt und die Schätzungen korrekt sind, liefert diese Methode die kleinste Varianz für alle unverzerrten Schätzer. Sie können eine Methode für die A-priori-Gewichtung der Zufallseffekte auswählen.
- Mit **ANOVA** (Varianzanalyse) werden unverzerrte Schätzungen unter Verwendung von Quadratsummen vom Typ I oder III für jeden Effekt berechnet. Die ANOVA-Methode liefert manchmal negative Varianzschätzungen. Dies kann auf ein falsches Modell, eine ungeeignete Schätzmethode oder eine ungenügende Anzahl von Daten hindeuten.
- Mit **Maximum Likelihood** (ML) werden unter Verwendung einer Iterationsmethode Schätzungen berechnet, die am ehesten den tatsächlich beobachteten Daten entsprechen. Diese Schätzungen können jedoch verzerrt sein. Diese Methode ist asymptotisch normalverteilt. ML- und REML-Schätzungen sind translationsinvariant. Bei dieser Methode werden die für die Schätzung der festen Effekte verwendeten Freiheitsgrade nicht berücksichtigt.
- Mit Schätzungen der **eingeschränkten Maximum Likelihood** (REML) werden die ANOVA-Schätzungen für viele (wenn nicht sogar alle) Fälle von ausgewogenen Daten reduziert. Da diese Methode für die festen Effekte angepasst ist, sind die Standardfehler normalerweise kleiner als bei der ML-Methode. Bei dieser Methode werden die für die Schätzung der festen Effekte verwendeten Freiheitsgrade berücksichtigt.

**A-priori-Annahmen für zufällige Effekte.** Gleich geht davon aus, dass alle Zufallseffekte und der Residuenterm den gleichen Einfluss auf die Beobachtung haben. Das **Nullschema** entspricht der Annahme von Nullvarianzen für die Zufallseffekte. Diese Optionen sind nur für die MINQUE-Methode verfügbar.

**Quadratsumme.** Quadratsummen vom **Typ I** werden für das hierarchische Modell angewandt (dies wird oft in der Literatur zu Varianzkomponenten verwendet). Wenn Sie **Typ III**, die Standardeinstellung in GLM, wählen, können Sie die Varianzschätzungen in der Prozedur "GLM - Univariat" für das Testen von Hypothesen mit Quadratsummen vom Typ III verwenden. Diese Optionen sind nur für die ANOVA-Methode verfügbar.

**Kriterien.** Sie können das Konvergenzkriterium und die maximale Anzahl der Iterationen angeben. Diese Optionen sind nur für die ML- und die REML-Methode verfügbar.

**Anzeigen.** Für die ANOVA-Methode können Sie Quadratsummen und erwartete Mittel der Quadrate anzeigen lassen. Wenn Sie die Methoden **Maximum Likelihood** oder **Eingeschränkte Maximum Likelihood** ausgewählt haben, können Sie den Iterationsverlauf anzeigen lassen.

## Quadratsummen (Varianzkomponenten)

Für das Modell können Sie einen Typ von Quadratsumme auswählen. Typ III wird am häufigsten verwendet und ist die Standardeinstellung.

**Typ I.** Diese Methode ist auch als die Methode der hierarchischen Zerlegung der Quadratsummen bekannt. Jeder Term wird nur für den Vorläuferterm im Modell korrigiert. Die Methode mit Quadratsummen vom Typ I wird gewöhnlich in den folgenden Fällen verwendet:

- Ein ausgewogenes ANOVA-Modell, in dem alle Haupteffekte vor den Interaktionseffekten 1. Ordnung festgelegt werden, alle Interaktionseffekte 1. Ordnung wiederum vor den Interaktionseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme niedrigerer Ordnung vor den Termen höherer Ordnung festgelegt werden.
- Ein rein verschachteltes Modell, in welchem der zuerst bestimmte Effekt in dem als zweiten bestimmten Effekt verschachtelt ist, der zweite Effekt wiederum im dritten und so weiter. (Diese Form der Verschachtelung kann nur durch Verwendung der Befehlssprache erreicht werden.)

**Typ III.** Der Standardparameter. Bei dieser Methode werden die Quadratsummen eines Effekts im Design als Quadratsummen orthogonal zu den Effekten (sofern vorhanden), die den Effekt enthalten, und mit Bereinigung um alle anderen Effekte, die diesen Effekt nicht enthalten, berechnet. Der große Vorteil der Quadratsummen vom Typ III ist, dass sie invariant bezüglich der Zellenhäufigkeiten sind, solange die allgemeine Form der Schätzbarkeit konstant bleibt. Daher wird dieser Typ oft für unausgewogene Modelle ohne fehlende Zellen als geeignet angesehen. In einem faktoriellen Design ohne fehlende Zellen ist diese Methode äquivalent zu der Methode der gewichteten Mittelwertquadrate nach Yates. Die Methode mit Quadratsummen vom Typ III wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I aufgeführten Modelle.
- Alle ausgewogenen oder unausgewogenen Modelle ohne leere Zellen.

## Varianzkomponenten: In neuer Datei speichern

Sie können einige Ergebnisse dieser Prozedur in einer neuen IBM SPSS Statistics-Datendatei speichern.

**Schätzungen der Varianzkomponenten.** Hiermit können Sie die Schätzungen der Varianzkomponenten und Beschriftungen für die Schätzungen in einer Datendatei oder in einem Dataset speichern. Diese können bei der Berechnung weiterer Statistiken oder für weitere Analysen in den GLM-Prozeduren verwendet werden. Hierbei kann es sich zum Beispiel um die Berechnung von Konfidenzintervallen oder um das Testen von Hypothesen handeln.

**Kovariation der Komponenten.** Hiermit können Sie eine Varianz-Kovarianz-Matrix oder eine Korrelationsmatrix in einer Datendatei oder in einem Dataset speichern. Dies ist nur verfügbar, wenn Sie **Maximum Likelihood** oder **Eingeschränkte Maximum Likelihood** ausgewählt haben.

**Ziel der erstellten Variablen.** Hiermit können Sie den Namen eines Datasets oder einer externen Datei festlegen, welches bzw. welche die Schätzungen der Varianzkomponenten und/oder die Matrix enthält. Datasets sind für die anschließende Verwendung in der gleichen Sitzung verfügbar, werden jedoch nicht als Dateien gespeichert, sofern Sie diese nicht ausdrücklich vor dem Beenden der Sitzung speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen. .

Sie können den Befehl MATRIX zum Auslesen der gewünschten Daten aus der Datendatei verwenden und anschließend Konfidenzintervalle berechnen oder Tests durchführen.

## Zusätzliche Funktionen beim Befehl VARCOMP

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl DESIGN können Sie verschachtelte Effekte im Design festlegen.
- Mit dem Unterbefehl MISSING können Sie benutzerdefiniert fehlende Werte aufnehmen.
- Mit dem Unterbefehl CRITERIA können Sie EPS-Kriterien angeben.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Lineare gemischte Modelle

Die Prozedur "Lineare gemischte Modelle" erweitert das allgemeine lineare Modell, indem sie zulässt, dass die Daten korrelierte und nicht konstante Variabilität aufweisen können. Das gemischte lineare Modell gibt Ihnen daher die Flexibilität, nicht nur die Mittelwerte der Daten, sondern auch ihre Varianzen und Kovarianzen zu analysieren.

Die Prozedur "Lineare gemischte Modelle" ist außerdem ein flexibles Tool zum Anpassen von anderen Modellen, die als gemischte lineare Modelle formuliert werden können. Dazu gehören Mehrebenenmodelle, hierarchische lineare Modelle und Zufallskoeffizientenmodelle.

### Beispiel

Eine Lebensmittelkette möchte die Effekte von verschiedenen Gutscheinen auf das Kaufverhalten von Kunden herausfinden. Nach der Entnahme einer Zufallsstichprobe ihrer Stammkunden werden 10 Wochen lang die Ausgaben jedes Kunden verfolgt. In jeder Woche wird den Kunden ein anderer Gutschein zugeschickt. Die Prozedur "Lineare gemischte Modelle" wird verwendet, um den Effekt der verschiedenen Gutscheine auf das Kaufverhalten zu schätzen und dabei für jedes Subjekt die Korrelation aufgrund von Beobachtungswiederholungen in den 10 Wochen anzupassen.

### Methoden

Schätzung der Maximum Likelihood (ML) und der eingeschränkten Maximum Likelihood (REML).

### Statistik

Deskriptive Statistiken: Stichprobengrößen, Mittelwerte und Standardabweichungen von abhängigen Variablen und Kovariaten für jede einzelne Kombination von Faktorstufen. Informationen zu Faktorstufe: sortierte Werte der Stufen für jeden Faktor und ihre Häufigkeiten. Darüber hinaus Parameterschätzungen und Konfidenzintervalle für feste Effekte, Wald-Tests und Konfidenzintervalle für Parameter von Kovarianzmatrizen. Zum Bewerten von verschiedenen Hypothesen können Quadratsummen vom Typ I und Typ III verwendet werden. Die Voreinstellung sieht den Typ III vor.

## Erläuterungen der Daten für lineare gemischte Modelle

### Daten

Die abhängige Variable muss quantitativ sein. Faktoren müssen kategorial sein und können numerische Werte oder Zeichenfolgewerte besitzen. Kovariaten und die Gewichtungvariable müssen quantitativ sein. Für Subjekte und Wiederholungsvariablen sind beliebige Typen möglich.

### Voraussetzungen

Es wird angenommen, dass die abhängige Variable sich linear zu den festen Faktoren, Zufallsfaktoren und Kovariaten verhält. Die festen Effekte analysieren den Mittelwert der abhängigen Variablen. Die Zufallseffekte analysieren die Kovarianzstruktur der abhängigen Variablen. Für jeden Zufallseffekt wird eine separate Kovarianzmatrix berechnet, da alle zufälligen Effekte als unabhängig voneinander betrachtet werden. Dennoch kann zwischen Modelltermen, die für denselben Zufallseffekt angegeben wurden, eine Korrelation bestehen. Die Messwiederholungen analysieren die Kovarianzstruktur der Residuen. Für die abhängige Variable wird außerdem angenommen, dass sie aus einer Normalverteilung stammt.

### Verwandte Prozeduren

Verwenden Sie die Prozedur "Explorative Datenanalyse", um die Daten vor der Durchführung der Analyse zu untersuchen. Wenn Sie keine korrelierte oder nicht konstante Variabilität vermuten, können Sie die Prozedur "GLM - Univariat" oder die Prozedur "GLM - Messwiederholung" verwenden. Darüber hinaus können Sie auch die Prozedur "Varianzkomponentenanalyse" verwenden, wenn die Zufallseffekte eine Kovarianzstruktur mit Varianzkomponenten aufweisen und keine Messwiederholungen vorhanden sind.

## Durchführen einer Analyse mit linearen gemischten Modellen

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Gemischte Modelle > Linear ...**

2. Wählen Sie optional mindestens eine Subjektvariable aus.

3. Wählen Sie optional mindestens eine Wiederholungsvariable aus. Wenn Wiederholungsvariablen definiert sind, wählen Sie aus der Dropdown-Liste einen **Kovarianztyp bei Messwiederholung** aus.

4. Wählen Sie optional mindestens eine Kronecker-Messvariable aus.

5. Wählen Sie optional eine Kovarianzstruktur der Residuen aus.

6. Klicken Sie auf **Weiter**.

7. Wählen Sie eine abhängige Variable aus.
  8. Wählen Sie mindestens einen Faktor oder eine Kovariate aus.
  9. Klicken Sie auf **Fest** oder **Zufällig** und geben Sie mindestens ein Modell mit festen oder zufälligen Effekten an.
- Wählen Sie optional eine GewichtungsvARIABLE aus.

## Lineare gemischte Modelle: Subjekte und Wiederholungen

In diesem Dialogfeld können Sie Variablen auswählen, die Subjekte, Beobachtungswiederholungen und Kronecker-Maße definieren, und eine Kovarianzstruktur für die Residuen angeben.

### Subjekte

Ein Subjekt ist eine Beobachtungseinheit, die als unabhängig von anderen Subjekten betrachtet werden kann. Die Blutdruckmessungen eines Patienten in einer medizinischen Studie können beispielsweise als unabhängig von den Messungen anderer Patienten angesehen werden. Die Definition von Subjekten ist vor allem dann wichtig, wenn für jedes Subjekt Messwiederholungen durchgeführt werden und Sie die Korrelation zwischen diesen Beobachtungen analysieren möchten. So ist beispielsweise zu erwarten, dass Blutdruckmessungen bei einem bestimmten Patienten bei aufeinander folgenden Arztbesuchen miteinander korrelieren.

Subjekte können auch durch die Faktorstufenkombination von verschiedenen Variablen definiert werden. Beispielsweise können Sie *Geschlecht* und *Alterskategorie* als Subjektvariablen angeben, um die Hypothese zu untersuchen, dass *Männer über 65* einander ähneln, aber unabhängig von *Männern unter 65* und *Frauen* sind.

Alle in der Liste **Subjekte** angegebenen Variablen werden verwendet, um Subjekte für die Kovarianzstruktur der Residuen zu definieren. Sie können einige oder alle Variablen verwenden, um Subjekte für die Kovarianzstruktur der Zufallseffekte zu definieren.

### Wiederholt

Die in dieser Liste angegebenen Variablen werden verwendet, um Beobachtungswiederholungen zu bezeichnen. So können beispielsweise mit einer einzigen Variablen für *Woche* alle 10 Wochen der Beobachtungen in einer medizinischen Studie bezeichnet werden oder die Variablen *Monat* und *Tag* können gemeinsam verwendet werden, um tägliche Beobachtungen im Verlauf eines Jahres zu bezeichnen.

### Kovarianztyp bei Messwiederholung

Hiermit wird die Kovarianzstruktur für die Residuen angegeben. Die verfügbaren Strukturen lauten folgendermaßen:

- Ante-Dependenz: 1. Ordnung
- AR(1)
- Direktes Produkt, AR1 (UN\_AR1)
- Direktes Produkt, unstrukturiert (UN\_UN)
- Direktprodukt, zusammengesetzte Symmetrie (UN\_CS)
- AR(1): Heterogen
- ARMA(1,1)
- Zusammengesetzt symmetrisch (ZS)
- Zusammengesetzt symmetrisch: Korrelationsmetrik
- Zusammengesetzt symmetrisch: heterogen
- Diagonal
- Faktoranalytisch: 1. Ordnung
- Faktoranalytisch: 1. Ordnung, heterogen
- Huynh-Feldt (HF)
- Skalierte Identität

- Toeplitz
- Toeplitz: heterogen
- Unstrukturiert (UN)
- Unstrukturiert: Korrelationsmetrik
- Räumlich: Exponent
- Räumlich: exponentiell
- Räumlich: gaußsch
- Räumlich: linear
- Räumlich: linear logistisch
- Räumlich: kugelförmig

### **Kronecker-Messungen**

Wählen Sie Variablen aus, die die Subjektstruktur für Kronecker-Kovarianz-Messungen angeben und bestimmen, wie die Messfehler korreliert werden. Das Feld ist nur verfügbar, wenn eine der folgenden Optionen unter **Kovarianztyp bei wiederholter Kovarianz** ausgewählt ist:

- Direktes Produkt, AR1 (UN\_AR1)
- Direktes Produkt, unstrukturiert (UN\_UN)
- Direktprodukt, zusammengesetzte Symmetrie (UN\_CS)

### **Koordinaten für räumliche Kovarianz**

Die Variablen in dieser Liste geben die Koordinaten der Beobachtungswiederholungen an, wenn einer der räumlichen Kovarianztypen als Kovarianztyp bei Messwiederholung ausgewählt wird.

Weitere Informationen finden Sie im Thema „Kovarianzstrukturen“ auf Seite 107.

## **Lineare gemischte Modelle: Feste Effekte**

**Feste Effekte.** Es gibt kein Standardmodell. Daher müssen Sie die festen Effekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme erstellen.

**Konstanter Term einschließen.** Der konstante Term wird für gewöhnlich in das Modell aufgenommen. Wenn anzunehmen ist, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

**Quadratsumme.** Die Methode zum Berechnen der Quadratsummen. Für Modelle ohne fehlende Zellen wird meistens die Methode vom Typ III verwendet.

### **Erstellen von nicht verschachtelten Termen**

Für die ausgewählten Faktoren und Kovariaten:

**Mehrfaktoriell.** Erstellt alle möglichen Interaktionen und Haupteffekte der ausgewählten Variablen. Dies ist der Standardwert.

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

**Haupteffekte.** Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

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

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

**Alle 4-Wege.** Erstellt alle möglichen Vierwegeinteraktionen der ausgewählten Variablen.

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

## Erstellen von verschachtelten Termen

In dieser Prozedur können Sie verschachtelte Terme für Ihr Modell erstellen. Verschachtelte Terme sind nützlich, um den Effekt eines Faktors oder einer Kovariaten zu modellieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. Eine Lebensmittelkette kann beispielsweise das Kaufverhalten ihrer Kunden in mehreren Filialen untersuchen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

Darüber hinaus können Sie Interaktionseffekte einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

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

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

## Quadratsumme

Für das Modell können Sie einen Typ von Quadratsumme auswählen. Typ III wird am häufigsten verwendet und ist die Standardeinstellung.

**Typ I.** Diese Methode ist auch als die Methode der hierarchischen Zerlegung der Quadratsummen bekannt. Jeder Term wird nur für den Vorläuferterm im Modell angepasst. Quadratsummen vom Typ I werden gewöhnlich in den folgenden Situationen verwendet:

- Ein ausgewogenes ANOVA-Modell, in dem alle Haupteffekte vor den Interaktionseffekten 1. Ordnung festgelegt werden, alle Interaktionseffekte 1. Ordnung wiederum vor den Interaktionseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme niedrigerer Ordnung vor den Termen höherer Ordnung festgelegt werden.
- Ein rein verschachteltes Modell, in welchem der zuerst bestimmte Effekt in dem als zweiten bestimmten Effekt verschachtelt ist, der zweite Effekt wiederum im dritten und so weiter. (Diese Form der Verschachtelung kann nur durch Verwendung der Befehlssprache erreicht werden.)

**Typ III.** Der Standardparameter. Bei dieser Methode werden die Quadratsummen eines Effekts im Design als Quadratsummen orthogonal zu den Effekten (sofern vorhanden), die den Effekt enthalten, und mit Bereinigung um alle anderen Effekte, die diesen Effekt nicht enthalten, berechnet. Der große Vorteil der Quadratsummen vom Typ III ist, dass sie invariant bezüglich der Zellenhäufigkeiten sind, solange die allgemeine Form der Schätzbarkeit konstant bleibt. Daher wird dieser Typ von Quadratsumme oft für unausgewogene Modelle ohne fehlende Zellen als geeignet angesehen. In einem faktoriellen Design ohne fehlende Zellen ist diese Methode äquivalent zu der Methode der gewichteten Mittelwertquadrate nach Yates. Die Methode mit Quadratsummen vom Typ III wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I aufgeführten Modelle.
- Alle ausgewogenen oder unausgewogenen Modelle ohne leere Zellen.

## Lineare gemischte Modelle: Zufällige Effekte

**Kovarianztyp.** Hiermit können Sie die Kovarianzstruktur für das Modell mit Zufallseffekten angeben. Für jeden Zufallseffekt wird eine separate Kovarianzmatrix geschätzt. Die verfügbaren Strukturen lauten folgendermaßen:

- Ante-Dependenz: 1. Ordnung
- AR(1)
- AR(1): Heterogen

- ARMA(1,1)
- Zusammengesetzt symmetrisch
- Zusammengesetzt symmetrisch: Korrelationsmetrik
- Zusammengesetzt symmetrisch: heterogen
- Diagonal
- Faktoranalytisch: 1. Ordnung
- Faktoranalytisch: 1. Ordnung, heterogen
- Huynh-Feldt
- Skalierte Identität
- Toeplitz
- Toeplitz: heterogen
- Unstrukturiert
- Unstrukturiert: Korrelationsmetrik
- Varianzkomponenten

Weitere Informationen finden Sie in „Kovarianzstrukturen“ auf Seite 107.

**Zufällige Effekte.** Es gibt kein Standardmodell. Daher müssen Sie die zufälligen Effekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme erstellen. Darüber hinaus können Sie einen konstanten Term in das Modell mit Zufallseffekten einschließen.

Sie können mehrere Modelle mit Zufallseffekten angeben. Klicken Sie nach der Erstellung des ersten Modells auf **Weiter**, um ein weiteres Modell zu erstellen. Klicken Sie auf **Zurück**, um die vorhandenen Modelle rückwärts zu durchblättern. Da alle Modelle mit zufälligen Effekten als unabhängig voneinander betrachtet werden, wird für jedes einzelne Modell eine separate Kovarianzmatrix berechnet. Dennoch kann zwischen Termen, die in demselben Modell mit Zufallseffekt angegeben wurden, eine Korrelation bestehen.

**Subjektgruppierungen.** Hier werden die Variablen aufgelistet, die zuvor im Dialogfeld "Variablen für Subjekte und Wiederholungen auswählen" angegeben wurden. Legen Sie die Subjekte für das Modell mit Zufallseffekten fest, indem Sie einige oder alle Variablen auswählen.

**Parametervorhersagen für diese Gruppe zufälliger Effekte anzeigen.** Gibt an, dass die Parameterschätzungen für zufällige Effekte angezeigt werden sollen.

## Lineare gemischte Modelle: Schätzung

### Methode

Wählen Sie die Schätzung für die Maximum Likelihood oder die eingeschränkte Maximum Likelihood aus.

### Freiheitsgrade

Stellt Optionen zur Verfügung, mit denen die Freiheitsgrade für alle Tests definiert werden können.

### Residualmethode

Die Residualmethode besitzt feste Freiheitsgrade für alle Tests. Sie ist hilfreich, wenn Ihre Stichprobe groß genug ist, die Daten ausgewogen sind oder das Modell einen einfacheren Kovarianztyp verwendet, z. B. "Skalierte Identität" oder "Diagonal".

### Satterthwaite-Approximation

Die Satterthwaite-Methode besitzt feldbasierte Freiheitsgrade über die Tests hinweg. Dies ist hilfreich, wenn Ihre Stichprobe klein ist oder die Daten unausgewogen sind oder das Modell einen komplizierten Kovarianztyp verwendet (zum Beispiel "Unstrukturiert").

### Kenward-Roger-Approximation

Die Kenward-Roger-Methode ermöglicht eine genauere Schätzung bei kleineren Stichproben für die Varianz-Kovarianz der Parameter mit festen Effekten und die ungefähren für den Nenner gültigen Freiheitsgrade in t-Tests und F-Tests. Die Methode führt einen Skalierungsfaktor für die

F-Statistik ein und schätzt diese und die für den Nenner gültigen Freiheitsgrade mithilfe einer Taylor-Serienerweiterung für die geschätzte zufällige Struktur innerhalb der Daten.

**Anmerkung:** Die Kenward-Roger-Methode wird in der modellbasierten Kovarianz verwendet (statt in der robusten Kovarianz). Wenn sowohl die Kenward-Roger-Methode als auch die robuste Kovarianz ausgewählt sind, wird die Kenward-Roger-Methode auf die modellbasierte Kovarianz angewendet. In diesem Fall wird folgende Warnung angezeigt: "Da die Kenward-Roger-Methode ausgewählt ist, wird die Methode der robusten Kovarianz in die Methode der modellbasierten Kovarianz geändert."

### Iterationen

Die folgenden Optionen stehen zur Verfügung:

#### Maximalzahl der Iterationen

Geben Sie eine nicht negative Ganzzahl an.

#### Maximalzahl für Schritthalbierungen

Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritthalbung erreicht ist. Geben Sie eine positive Ganzzahl an.

#### Iterationsverlauf ausgeben alle $n$ Schritte

Zeigt eine Tabelle an, die den Log Likelihood-Funktionswert und Parameterschätzungen bei jeder  $N$ -Iteration enthält, die mit der 0<sup>th</sup>-Iteration beginnt (die ursprünglichen Schätzungen). Wenn Sie den Iterationsverlauf drucken, wird die letzte Iteration stets unabhängig vom Wert für  $n$  ausgegeben.

### Log-Likelihood-Konvergenz

Es wird angenommen, dass eine Konvergenz vorliegt, wenn die absolute oder relative Änderung in der Log-Likelihood-Funktion kleiner als der angegebene Wert ist, der nicht negativ sein darf. Wenn der angegebene Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

### Parameterkonvergenz

Es wird angenommen, dass eine Konvergenz vorliegt, wenn die maximale absolute oder maximale relative Änderung in den Parameterschätzungen kleiner als der angegebene Wert ist, der nicht negativ sein darf. Wenn der angegebene Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

### Konvergenz der Hesse-Matrix

Für die Spezifikation **Absolut** wird angenommen, dass eine Konvergenz vorliegt, wenn eine Statistik auf der Basis der Hesse-Matrix kleiner als der angegebene Wert ist. Für die Spezifikation **Relativ** wird angenommen, dass eine Konvergenz vorliegt, wenn die Statistik kleiner als das Produkt aus dem angegebenen Wert und dem absoluten Wert der Log-Likelihood ist. Wenn der angegebene Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

### Maximale Scoring-Schritte

Anforderungen, den Scoring-Algorithmus von Fisher bis zur Iterationsnummer  $N$  zu verwenden. Geben Sie eine nicht negative Ganzzahl an.

### Singularitätstoleranz

Dieser Wert wird als Toleranz bei der Prüfung auf Singularität verwendet. Geben Sie einen positiven Wert ein.

## Lineare gemischte Modelle: Statistik

**Auswertungsstatistik.** Es werden die folgenden Tabellen erstellt:

- **Deskriptive Statistiken.** Zeigt die Stichprobengrößen, Mittelwerte und Standardabweichungen der abhängigen Variablen und (falls angegeben) der Kovariaten an. Diese Statistik wird für jede einzelne Kombination von Faktorstufen angezeigt.
- **Zusammenfassung der Fallverarbeitung.** Zeigt die sortierten Werte der Faktoren, der Variablen für Messwiederholungen, der Subjekte für Messwiederholungen und der Subjekte mit Zufallseffekten sowie deren Häufigkeiten an.

**Modellstatistik.** Es werden die folgenden Tabellen erstellt:

- **Parameterschätzungen für feste Effekte.** Zeigt die Parameterschätzungen für feste Effekten sowie deren approximativen Standardfehler an.
- **Tests auf Kovarianzparameter.** Zeigt die asymptotischen Standardfehler und Wald-Tests für die Kovarianzparameter an.
- **Korrelationen der Parameterschätzungen.** Zeigt die asymptotische Korrelationsmatrix der Parameterschätzungen mit festen Effekten an.
- **Kovarianzen der Parameterschätzungen.** Zeigt die asymptotische Kovarianzmatrix der Parameterschätzungen mit festen Effekten an.
- **Kovarianzen der zufälligen Effekte.** Zeigt die geschätzte Kovarianzmatrix der Zufallseffekte an. Diese Option ist nur verfügbar, wenn mindestens ein Zufallseffekt angegeben wurde. Wenn eine Subjektvariable für einen Zufallseffekt angegeben ist, wird der gemeinsame Block angezeigt.
- **Kovarianzen der Residuen.** Zeigt die geschätzte Kovarianzmatrix der Residuen an. Diese Option ist nur verfügbar, wenn eine Wiederholungsvariable angegeben wurde. Wenn eine Subjektvariable angegeben ist, wird der gemeinsame Block angezeigt.
- **Kontrastkoeffizientenmatrix.** Diese Option zeigt die schätzbaren Funktionen an, die zum Testen der festen Effekte und der benutzerdefinierten Hypothesen verwendet werden.

**Konfidenzintervall.** Dieser Wert wird immer dann verwendet, wenn ein Konfidenzintervall erstellt wird. Geben Sie einen Wert größer-gleich 0 und kleiner als 100 ein. Der Standardwert ist 95.

## Lineare gemischte Modelle: Geschätzte Randmittel

**Geschätzte Randmittel von angepassten Modellen.** Mit dieser Gruppe können Sie vom Modell vorhergesagte geschätzte Randmittel der abhängigen Variable in den Zellen und die dazugehörigen Standardfehler für die festgelegten Faktoren anfordern. Außerdem können Sie einen Vergleich der Faktorstufen von Haupteffekten veranlassen.

- **Faktoren und Interaktionen zwischen Faktoren.** Diese Liste enthält Faktoren und Interaktionen zwischen Faktoren, die im Dialogfeld "Fest" angegeben wurden, sowie einen OVERALL-Term. Aus Kovarianzen erstellte Modellterme sind aus dieser Liste ausgeschlossen.
- **Mittelwerte anzeigen für.** Dieses Verfahren berechnet die geschätzten Randmittel für die Faktoren und Interaktionen zwischen Faktoren, die für diese Liste ausgewählt wurden. Wenn OVERALL ausgewählt wurde, werden die geschätzten Randmittel der abhängigen Variablen angezeigt, wobei diese über alle Faktoren zusammengefasst werden. Beachten Sie, dass die ausgewählten Faktoren oder Interaktionen zwischen Faktoren so lange ausgewählt bleiben, bis eine zugeordnete Variable im Hauptdialogfeld aus der Liste "Faktoren" entfernt wird.
- **Haupteffekte vergleichen.** Mit dieser Option können Sie paarweise Vergleiche von Stufen ausgewählter Haupteffekte anfordern. Mit der Anpassung des Konfidenzintervalls können Sie Konfidenzintervalle und Signifikanzwerte anpassen, um Mehrfachvergleiche zu berücksichtigen. Die verfügbaren Methoden sind LSD (keine Anpassung), Bonferroni und Sidak. Abschließend können Sie für jeden Faktor eine Referenzkategorie auswählen, anhand derer Vergleiche vorgenommen werden. Wenn keine Referenzkategorie ausgewählt wurde, werden alle paarweisen Vergleiche konstruiert. Als Referenzkategorie können Sie entweder die erste, die letzte oder eine benutzerdefinierte Kategorie angeben (letztere durch Eingabe von deren Wert).

## Lineare gemischte Modelle: Speichern

In diesem Dialogfeld können Sie mehrere Modellergebnisse in der Arbeitsdatei speichern.

**Feste vorhergesagte Werte.** Speichert Variablen, die sich auf den Mittelwert der Regression ohne Effekte beziehen.

- **Vorhergesagte Werte.** Der Mittelwert der Regression ohne die Zufallseffekte.
- **Standardfehler.** Die Standardfehler der Schätzungen.
- **Freiheitsgrade.** Die den Schätzungen zugeordneten Freiheitsgrade.

**Vorhergesagte Werte & Residuen.** Speichert Variablen im Zusammenhang mit dem an das Modell angepassten Wert.

- **Vorhergesagte Werte.** Der an das Modell angepasste Wert.
- **Standardfehler.** Die Standardfehler der Schätzungen.
- **Freiheitsgrade.** Die den Schätzungen zugeordneten Freiheitsgrade.
- **Residuen.** Der Datenwert abzüglich des vorhergesagten Werts.

## Lineare gemischte Modelle-Exportieren

In diesem Dialogfeld können Sie den Inhalt der Ausgabetafeln von EBLUPS (Empirische Beste Lineare Unverzerrte Vorhersagen) in Datasets oder .sav -Dateien exportieren. Die Schaltfläche Exportieren im Dialogfeld "Lineare gemischte Modelle" ist aktiviert, wenn mindestens ein Zufallseffekt über das Dialogfeld "Zufällig" angegeben und das Feld "Parametervorhersagen für diesen Block anzeigen" aktiviert ist.

### EBLUPS exportieren

Wählen Sie ein Ziel aus: **Dataset** oder **Datendatei**

Geben Sie optional einen Namen an.

Wenn mehrere zufällige Effekte mit erzeugten EBLUPS angegeben wurden, inaktivieren Sie das Kontrollkästchen, damit jede Ergebnistabelle in einem separaten Dataset oder einer separaten Datei angezeigt wird.

## Zusätzliche Funktionen beim Befehl MIXED

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl TEST können Sie Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl MISSING können Sie benutzerdefiniert fehlende Werte aufnehmen.
- Mit dem Schlüsselwort WITH im Unterbefehl EMMEANS können Sie geschätzte Randmittel für angegebene Werte von Kovariaten berechnen.
- Mit dem Unterbefehl EMMEANS können Sie einfache Haupteffekte von Interaktionen vergleichen.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Verallgemeinerte lineare Modelle

Das verallgemeinerte lineare Modell erweitert das allgemeine lineare Modell so, dass die abhängige Variable über eine angegebene Verknüpfungsfunktion in linearem Zusammenhang zu den Faktoren und Kovariaten steht. Außerdem ist es mit diesem Modell möglich, dass die abhängige Variable eine von der Normalverteilung abweichende Verteilung aufweist. Es deckt durch seine sehr allgemein gehaltene Modellformulierung häufig verwendete statistische Modelle ab, wie beispielsweise die lineare Regression für normalverteilte Antworten, logistische Modelle für binäre Daten und loglineare Modelle für Häufigkeitsdaten, Modelle vom Typ "Log-Log komplementär" für intervallzensierte Überlebensdaten sowie viele andere statistische Modelle.

**Beispiele.** Eine Reederei kann verallgemeinerte lineare Modelle verwenden, um eine Poisson-Regression an die Anzahl von Beschädigungen für mehrere Schiffstypen anzupassen, die in verschiedenen Zeiträumen gebaut wurden. Anhand des daraus hervorgehenden Modells kann bestimmt werden, welche Schiffstypen besonders anfällig für Schäden sind.

Ein KFZ-Versicherungsunternehmen kann verallgemeinerte lineare Modelle verwenden, um eine Gamma-Regression an die Schadensersatzansprüche für Autos anzupassen. Anhand des daraus hervorgehenden Modells können die Faktoren bestimmt werden, die am meisten zur Anspruchshöhe beitragen.

Mediziner können mithilfe von verallgemeinerten linearen Modellen eine komplementäre Log-Log-Regression an intervallzensierte Überlebensdaten anpassen, um die Dauer bis zum Wiederauftreten einer Erkrankung vorherzusagen.

Erläuterungen der Daten für verallgemeinerte lineare Modelle

**Daten.** Die Antwort kann metrisch (Skala), eine Anzahl, binär oder vom Typ "Ereignisse in Versuchen" sein. Es wird davon ausgegangen, dass Faktoren kategorial sind. Es wird davon ausgegangen, dass Kovariaten, Skalengewichtung und Offset metrisch (Skala) sind.

**Annahmen.** Fälle werden als unabhängige Beobachtungen betrachtet.

So erstellen Sie ein verallgemeinertes lineares Modell:

Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Verallgemeinerte lineare Modelle > Verallgemeinerte lineare Modelle...**

1. Geben Sie eine Verteilung und eine Verknüpfungsfunktion an (unten finden Sie Einzelheiten zu den verschiedenen Optionen).
2. Wählen Sie auf der Registerkarte Abhängige Variable eine abhängige Variable aus.
3. Wählen Sie auf der Registerkarte Prädiktoren die Faktoren und Kovariaten aus, die zur Vorhersage der abhängigen Variablen verwendet werden sollen.
4. Geben Sie auf der Registerkarte Modell mithilfe der ausgewählten Faktoren und Kovariaten Modelleffekte an.

Auf der Registerkarte "Typ des Modells" können Sie die Verteilung und die Verknüpfungsfunktion für Ihr Modell angeben, die Schnellverfahren für mehrere häufig verwendete Modelle bieten, die durch den Antworttyp kategorisiert sind.

Modelltypen

**Metrische abhängige Variable.** Die folgenden Optionen stehen zur Verfügung:

- **Linear.** Gibt "Normal" als Verteilung und "Identität" als Verknüpfungsfunktion an.
- **Gamma mit Log-Verknüpfung.** Gibt "Gamma" als Verteilung und "Log" als Verknüpfungsfunktion an.

**Ordinale abhängige Variable.** Die folgenden Optionen stehen zur Verfügung:

- **Ordinal logistisch.** Gibt "Multinomial (ordinal)" als Verteilung und "Logit (kumulativ)" als Verknüpfungsfunktion an.
- **Ordinal Probit.** Gibt "Multinomial (ordinal)" als Verteilung und "Probit (kumulativ)" als Verknüpfungsfunktion an.

**Häufigkeiten.** Die folgenden Optionen stehen zur Verfügung:

- **Poisson loglinear.** Gibt "Poisson" als Verteilung und "Log" als Verknüpfungsfunktion an.
- **Negativ binomial mit Log-Verknüpfung.** Gibt "Negativ binomial" (mit einem Wert von 1 für den Hilfsparameter) als Verteilung und "Log" als Verknüpfungsfunktion an. Damit die Prozedur den Wert des Hilfsparameters schätzt, geben Sie ein benutzerdefiniertes Modell mit negativer binomialer Verteilung an und wählen Sie **Wert schätzen** in der Gruppe "Parameter".

**Binär oder Anzahl Ereignisse.** Die folgenden Optionen stehen zur Verfügung:

- **Binär logistisch.** Gibt "Binomial" als Verteilung und "Logit" als Verknüpfungsfunktion an.
- **Binär Probit.** Gibt "Binomial" als Verteilung und "Probit" als Verknüpfungsfunktion an.
- **Intervallzensiertes Überleben.** Gibt "Binomial" als Verteilung und "Log-Log komplementär" als Verknüpfungsfunktion an.

**Mischung.** Die folgenden Optionen stehen zur Verfügung:

- **Tweedie mit Verknüpfungsfunktion "Log".** Gibt "Tweedie" als Verteilung und "Log" als Verknüpfungsfunktion an.
- **Tweedie mit Verknüpfungsfunktion "Identität".** Gibt "Tweedie" als Verteilung und "Identität" als Verknüpfungsfunktion an.

**Benutzerdefiniert.** Geben Sie Ihre eigene Kombination von Verteilung und Verknüpfungsfunktion an.

Verteilung

Diese Auswahl gibt die Verteilung der abhängigen Variablen an. Die Möglichkeit einer anderen Verteilung als "Normal" und einer anderen Verknüpfungsfunktion als "Identität" ist die wichtigste Verbesserung des verallgemeinerten linearen Modells gegenüber dem allgemeinen linearen Modell. Es gibt zahlreiche mögliche Kombinationen aus Verteilung und Verknüpfungsfunktion und es können mehrere davon für das jeweils vorliegende Dataset geeignet sein. Daher können Sie sich in Ihrer Wahl durch theoretische Vorüberlegungen leiten lassen oder davon, welche Kombination am besten zu passen scheint.

- **Binomial.** Diese Verteilung ist nur für Variablen geeignet, die eine binäre Antwort oder eine Anzahl von Ereignissen repräsentieren.
- **Gamma.** Diese Verteilung eignet sich für Variablen mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Wenn ein Datenwert kleiner oder gleich 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Invers normal.** Diese Verteilung eignet sich für Variablen mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Wenn ein Datenwert kleiner oder gleich 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Negativ binomial.** Diese Verteilung lässt sich als Anzahl der Versuche betrachten, die erforderlich sind, um  $k$  Erfolge zu beobachten, und eignet sich für Variablen mit nicht negativen ganzzahligen Werten. Wenn ein Datenwert keine Ganzzahl oder kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet. Der Wert für den Hilfsparameter der negativen Binomialverteilung kann eine beliebige Zahl größer als oder gleich 0 sein. Sie können ihn auf einen festen Wert setzen oder ihn nach dem Verfahren schätzen lassen. Wenn der Hilfsparameter auf 0 gesetzt wird, ist die Verwendung dieser Verteilung äquivalent zur Verwendung der Poisson-Verteilung.
- **Normal.** Diese Option eignet sich für metrische Variablen, deren Werte eine symmetrische, glockenförmige Verteilung um einen Mittelwert aufweisen. Die abhängige Variable muss numerisch sein.
- **Poisson-Verteilung.** Diese Verteilung lässt sich als Anzahl der Vorkommen eines untersuchten Ereignisses in einem festen Zeitraum betrachten und eignet sich für Variablen mit nicht negativen ganzzahligen Werten. Wenn ein Datenwert keine Ganzzahl oder kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Tweedie.** Diese Verteilung eignet sich für Variablen, die durch Poisson-Mischungen von Gammaverteilungen repräsentiert werden können; die Verteilung ist dahingehend "gemischt", dass sie sowohl Eigenschaften von stetigen Verteilungen (nimmt nicht negative reelle Werte an) als auch von diskreten Verteilungen (positive Wahrscheinlichkeitsmasse an einem Einzelwert, 0) aufweist. Die abhängige Variable muss numerisch sein, mit Datenwerten größer oder gleich 0. Wenn ein Datenwert kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet. Der feste Wert des Parameters der Tweedie-Verteilung kann jede beliebige Zahl zwischen 1 und 2 sein.
- **Multinomial.** Diese Verteilung eignet sich für Variablen, die eine ordinale abhängige Variable repräsentieren. Bei der abhängigen Variablen kann es sich um eine numerische Variable oder eine Zeichenfolgevariable handeln; sie muss mindestens zwei verschiedene gültige Datenwerte aufweisen.

#### Verknüpfungsfunktionen

Die Verknüpfungsfunktion ist eine Transformation der abhängigen Variable, die eine Schätzung des Modells ermöglicht. Die folgenden Funktionen sind verfügbar:

- **Identität.**  $f(x)=x$ . Die abhängige Variable ist nicht transformiert. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Komplementäres Protokoll-Protokoll.**  $f(x)=\log(-\log(1-x))$ . Nur für die Binomialverteilung geeignet.
- **Kumulative Cauchit.**  $f(x) = \tan(\pi(x - 0,5))$ , angewendet auf die kumulative Wahrscheinlichkeit für jede Kategorie der Antwort. Nur für die Multinomialverteilung geeignet.
- **Kumulatives komplementäres Protokoll-Protokoll.**  $f(x) = \ln(-\ln(1-x))$ , angewendet auf die kumulative Wahrscheinlichkeit für jede Kategorie der Antwort. Nur für die Multinomialverteilung geeignet.
- **Logit (kumulativ).**  $f(x)=\ln(x / (1-x))$ , auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Nur für die Multinomialverteilung geeignet.
- **Kumulatives negatives Protokoll-Protokoll.**  $f(x) = -\ln(-\ln(x))$ , angewendet auf die kumulative Wahrscheinlichkeit für jede Kategorie der Antwort. Nur für die Multinomialverteilung geeignet.

- **Probit (kumulativ).**  $f(x)=\Phi^{-1}(x)$ , auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Dabei steht  $\Phi^{-1}$  für die umgekehrte kumulative Standardnormalverteilungsfunktion steht. Nur für die Multinomialverteilung geeignet.
- **Log.**  $f(x) = \log(x)$ . Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Protokollergänz.**  $f(x) = \log(1-x)$ . Nur für die Binomialverteilung geeignet.
- **Logit.**  $f(x) = \log(x/(1-x))$ . Nur für die Binomialverteilung geeignet.
- **Negatives Binomial.**  $f(x) = \log(x/(x+k^{-1}))$ . Hierbei steht  $k$  für den Nebenparameter der negativen Binomialverteilung. Nur für die negative Binomialverteilung geeignet.
- **Negatives Protokoll-Protokoll.**  $f(x) = -\log(-\log(x))$ . Nur für die Binomialverteilung geeignet.
- **-Odds Power.**  $f(x)=[(x/(1-x))^\alpha-1]/\alpha$ , wenn  $\alpha \neq 0$  ist.  $f(x)=\log(x)$ , wenn  $\alpha = 0$  ist.  $\alpha$  ist die erforderliche Zahlenangabe und muss eine reelle Zahl sein. Nur für die Binomialverteilung geeignet.
- **Probit.**  $f(x) = \Phi^{-1}(x)$ , wobei  $\Phi^{-1}$  die umgekehrte kumulative Standardnormalverteilungsfunktion ist. Nur für die Binomialverteilung geeignet.
- **Power.**  $f(x) = x^\alpha$ , wenn  $\alpha \neq 0$  ist.  $f(x)=\log(x)$ , wenn  $\alpha = 0$  ist.  $\alpha$  ist die erforderliche Zahlenangabe und muss eine reelle Zahl sein. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.

## Verallgemeinerte lineare Modelle: Antwort

In vielen Fällen können Sie einfach eine abhängige Variable angeben. Allerdings erfordern Variablen, die nur zwei Werte annehmen, und Antworten, die Ereignisse in Versuchen aufzeichnen, besondere Aufmerksamkeit.

- **Binäre Antwort.** Wenn die abhängige Variable nur zwei Werte annimmt, können Sie die Referenzkategorie für die Parameterschätzung angeben. Binäre Antwortvariablen können Zeichenfolgevariablen oder numerische Variablen sein.
- **Anzahl der Ereignisse, die in einem Set von Versuchen vorkommen.** Wenn die Antwort eine Anzahl von Ereignissen ist, die in einem Set von Versuchen vorkommen, enthält die abhängige Variable die Anzahl der Ereignisse und Sie können eine zusätzliche Variable auswählen, die die Anzahl der Versuche enthält. Wenn die Anzahl der Versuche über alle Subjekte gleich ist, können die Versuche alternativ auch über einen festen Wert angegeben werden. Die Anzahl der Versuche sollte größer oder gleich der Anzahl der Ereignisse für jeden Fall sein. Bei den Ereignissen sollte es sich um nicht negative Ganzzahlen und bei den Versuchen um positive Ganzzahlen handeln.

Bei ordinal-multinomialen Modellen können Sie die Kategoriereihenfolge der Antwort angeben: aufsteigend, absteigend bzw. Daten (die Reihenfolge "Daten" bedeutet, dass der erste in den Daten gefundene Wert die erste Kategorie definiert und der letzte gefundene Wert die letzte Kategorie).

**Skalengewichtung.** Der Skalenparameter ist ein geschätzter Modellparameter, der mit der Varianz der Antwort zusammenhängt. Die Skalengewichtungen sind "bekannte" Werte, die sich zwischen den einzelnen Beobachtungen unterscheiden können. Wenn die Skalengewichtungsvariable angegeben ist, wird der Skalenparameter, der mit der Varianz der Antwort zusammenhängt, für jede Beobachtung durch diese Variable geteilt. Fälle, deren Skalengewichtungswerte kleiner-gleich 0 sind oder fehlen, werden in der Analyse nicht verwendet.

## Verallgemeinerte lineare Modelle: Referenzkategorie

Bei binären Antworten können Sie die Referenzkategorie für die abhängige Variable auswählen. Dies kann sich auf bestimmte Ausgaben, wie beispielsweise Parameterschätzungen und gespeicherte Werte, auswirken, sollte jedoch nicht die Anpassungsgüte des Modells verändern. Beispiel: Angenommen, Ihre binäre Antwort nimmt die Werte 0 und 1 an:

- Standardmäßig verwendet die Prozedur die letzte Kategorie (die mit dem höchsten Wert), also 1, als Referenzkategorie. In dieser Situation wird anhand der im Modell gespeicherten Wahrscheinlichkeitswerte die Wahrscheinlichkeit geschätzt, dass ein bestimmter Fall den Wert 0 annimmt, und die Parameterschätzungen sollten als in Bezug zur Likelihood von Kategorie 0 betrachtet werden.

- Wenn Sie die erste Kategorie (die mit dem niedrigsten Wert), also 0, als Referenzkategorie angeben, wird anhand der im Modell gespeicherten Wahrscheinlichkeitswerte die Wahrscheinlichkeit geschätzt, dass ein bestimmter Fall den Wert 1 annimmt.
- Wenn Sie die benutzerdefinierte Kategorie angeben und für Ihre Variable Beschriftungen definiert sind, können Sie die Referenzkategorie durch Auswahl eines Werts aus der Liste festlegen. Dies kann nützlich sein, wenn Sie bei der Festlegung eines Modells nicht mehr wissen, wie genau eine bestimmte Variable codiert war.

## Verallgemeinerte lineare Modelle: Prädiktoren

Auf der Registerkarte "Prädiktoren" können Sie die Faktoren und Kovariaten angeben, die zum Erstellen der Modelleffekte und zur Angabe eines optionalen Offsets verwendet werden.

**Faktoren.** Faktoren sind kategoriale Prädiktoren; es kann sich dabei um numerische Variablen oder Zeichenfolgevariablen handeln.

**Kovariaten.** Kovariaten sind metrische Prädiktoren; sie müssen numerisch sein.

*Hinweis:* Wenn die Antwort binomial mit Binärformat ist, berechnet die Prozedur Abweichungs- und Chi-Quadrat-Anpassungsstatistiken anhand von Teilgesamtheiten, die auf der Kreuzklassifikation beobachteter Werte der ausgewählten Faktoren und Kovariaten beruhen. Sie sollten dasselbe Set an Prädiktoren während mehrerer Ausführungen der Prozedur beibehalten, um eine konsistente Anzahl an Teilgesamtheiten zu gewährleisten.

**Offset.** Der Term "Offset" ist ein "struktureller" Prädiktor. Sein Koeffizient wird nicht vom Modell geschätzt, sondern es wird davon ausgegangen, dass er den Wert 1 aufweist. Daher werden die Werte des Offsets einfach zum linearen Prädiktor des Ziels addiert. Dies ist besonders nützlich bei Poisson-Regressionsmodellen, bei denen die verschiedenen Fälle dem relevanten Ereignis unterschiedlich stark ausgesetzt sein können.

Beispielsweise gibt es bei der Modellierung der Unfallraten für einzelne Fahrer einen wichtigen Unterschied zwischen einem Fahrer, der in 3 Jahren Fahrpraxis einen Unfall verursacht hat und einem Fahrer, der in 25 Jahren einen Unfall verursacht hat. Die Anzahl der Unfälle kann als Poisson- oder negative Binomialantwort mit einer Protokollverknüpfung modelliert werden, wenn der natürliche Logarithmus der Fahrpraxis des Fahrers als Offset-Term eingeschlossen wird.

Andere Kombinationen der Verteilung und Verknüpfungstypen machen andere Transformationen der Offsetvariable erforderlich.

## Verallgemeinerte lineare Modelle: Optionen

Diese Optionen werden auf alle auf der Registerkarte "Prädiktoren" angegebenen Faktoren angewendet.

**Benutzerdefiniert fehlende Werte.** Faktoren müssen gültige Werte für einen Fall aufweisen, um in die Analyse aufgenommen zu werden. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Faktorvariablen als gültige Werte behandelt werden sollen.

**Kategoriereihenfolge.** Diese ist relevant für die Bestimmung der letzten Stufe eines Faktors, die einem redundanten Parameter im Schätzalgorithmus zugeordnet werden kann. Durch das Ändern der Kategoriereihenfolge können sich die Werte der Faktorstufeneffekte ändern, da diese Parameterschätzungen bezogen auf die letzte Stufe berechnet werden. Die Faktoren können in aufsteigender Reihenfolge vom niedrigsten zum höchsten Wert, in absteigender Reihenfolge vom höchsten zum niedrigsten Wert oder in "Datenreihenfolge" angeordnet werden. Dies bedeutet, dass der erste Wert, der in den Daten vorgefunden wird, die erste Kategorie definiert und der letzte eindeutige Wert, der vorgefunden wird, die letzte Kategorie.

## Verallgemeinerte lineare Modelle: Modell

### Modelleffekte angeben

Das Standardmodell ist "Nur konstanter Term". Daher müssen Sie die anderen Modelleffekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme erstellen.

## Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

### Haupteffekte

Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

### Interaktion

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

### Faktoriell

Erstellt alle möglichen Interaktionen und Haupteffekte der ausgewählten Variablen.

### Alle 2-Wege

Hiermit werden alle möglichen Zweibegeinteraktionen der ausgewählten Variablen erstellt.

### Alle 3-Wege

Hiermit werden alle möglichen Dreibegeinteraktionen der ausgewählten Variablen erzeugt.

### Alle 4-Wege

Erstellt alle möglichen Vierbegeinteraktionen der ausgewählten Variablen.

### Alle 5-Wege

Hiermit werden alle möglichen Fünfbegeinteraktionen der ausgewählten Variablen erzeugt.

## Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für Ihr Modell erstellen. Verschachtelte Terme sind nützlich, um den Effekt eines Faktors oder einer Kovariate zu modellieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. So kann beispielsweise eine Lebensmittelkette den Ausgabengewohnheiten ihrer Kunden an mehreren Filialstandorten folgen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

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

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

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

## Verallgemeinerte lineare Modelle: Schätzung

**Parameterschätzung.** Mit den Steuerelementen in dieser Gruppe können Sie Schätzmethoden festlegen und Anfangswerte für die Parameterschätzungen angeben.

- **Methode.** Sie können eine Methode für die Parameterschätzung auswählen. Sie haben die Wahl zwischen "Newton-Raphson", "Fisher-Bewertung" und einer Hybridmethode, bei der zuerst Iterationen des Fisher-Scorings durchgeführt werden und dann zur Methode "Newton-Raphson" gewechselt wird. Wenn während der Phase "Fisher-Scoring" der Hybridmethode Konvergenz erreicht wird, bevor die maximale Anzahl an Fisher-Iterationen erreicht wurde, fährt der Algorithmus mit der Newton-Raphson-Methode fort.
- **Skalenparametermethode.** Sie können eine Schätzmethode für den Skalenparameter auswählen. Bei der Maximum-Likelihood-Methode wird der Skalenparameter zusammen mit den Modelleffekten geschätzt. Beachten Sie, dass diese Option nicht gültig ist, wenn die Antwort eine negative Binomialverteilung, eine Poisson-Verteilung, eine Binomialverteilung oder eine multinomiale Verteilung aufweist. Die Optionen für die Abweichung und Pearson-Chi-Quadrat schätzen den Skalenparameter aus dem Wert der jeweiligen Statistik. Alternativ können Sie einen festen Wert für den Skalenparameter angeben.

- **Anfangswerte.** Die Prozedur berechnet automatisch Anfangswerte für Parameter. Alternativ können Sie Anfangswerte für die Parameterschätzungen angeben.

**Kovarianzmatrix.** Der modellbasierte Schätzer ist das Negative der verallgemeinerten Inversen der Hesse-Matrix. Der robuste Schätzer (auch Huber-/White-/Sandwich-Schätzer genannt) ist ein "korrigierter" modellbasierter Schätzer, der eine konsistente Schätzung der Kovarianz bietet, selbst wenn Varianz und Verknüpfungsfunktionen falsch angegeben wurden.

**Iterationen.** Die folgenden Optionen stehen zur Verfügung:

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nicht negative Ganzzahl an.
- **Maximalzahl für Schritthalbierung.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritthalbierung erreicht ist. Geben Sie eine positive Ganzzahl an.
- **Prüfen auf Datenpunkttrennung.** Mit dieser Option lassen Sie Tests durch den Algorithmus durchführen, mit denen sichergestellt wird, dass die Parameterschätzungen eindeutige Werte aufweisen. Eine Trennung wird vorgenommen, sobald ein Modell erzeugt werden kann, in dem alle Fälle fehlerfrei klassifiziert werden. Diese Option ist für multinomiale Antworten und binomiale Antworten mit Binärformat verfügbar.

**Konvergenzkriterien.** Die folgenden Optionen sind verfügbar:

- **Parameterkonvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzungen unter dem angegebenen (positiven) Wert liegt.
- **Log-Likelihood-Konvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei der Log-Likelihood-Funktion unter dem angegebenen (positiven) Wert liegt.
- **Konvergenz der Hesse-Matrix.** Für die Spezifikation 'Absolut' wird angenommen, dass eine Konvergenz vorliegt, wenn eine Statistik auf der Basis der Konvergenz der Hesse-Matrix kleiner als der angegebene positive Wert ist. Für die Spezifikation "Relativ" wird angenommen, dass eine Konvergenz vorliegt, wenn die Statistik kleiner als das Produkt aus dem angegebenen positiven Wert und dem absoluten Wert der Log-Likelihood ist.

**Toleranz für Prüfung auf Singularität.** Singuläre (bzw. nicht invertierbare) Matrizen weisen linear abhängige Spalten auf, die zu ernststen Problemen für den Schätzalgorithmus führen können. Auch annähernd singuläre Matrizen können zu schlechten Ergebnissen führen, daher behandelt die Prozedur eine Matrix, deren Determinante unter dem Toleranzwert liegt, als singulär. Geben Sie einen positiven Wert an.

## Verallgemeinerte lineare Modelle: Anfangswerte

Wenn Anfangswerte angegeben wurden, müssen diese für alle Parameter (auch redundante Parameter) im Modell angegeben werden. Im Dataset müssen die Variablen von links nach rechts folgende Reihenfolge aufweisen: *RowType\_*, *VarName\_*, *P1*, *P2* usw. Dabei sind *RowType\_* und *VarName\_* Zeichenfolgevariablen und *P1*, *P2* usw. numerische Variablen, die einer geordneten Liste der Parameter entsprechen.

- Die Anfangswerte werden in einem Datensatz angegeben, mit dem Wert *EST* für die Variable *RowType\_*. Die tatsächlichen Anfangswerte werden unter den Variablen *P1*, *P2* usw. angegeben. Die Prozedur ignoriert alle Datensätze, für die *RowType\_* einen anderen Wert als *EST* aufweist sowie alle Datensätze nach dem ersten Vorkommen von *RowType\_*, die gleich *EST* sind.
- Der konstante Term, sofern im Modell enthalten, bzw. die Schwellenwertparameter, falls das Modell eine Multinomialverteilung aufweist, müssen als erste Anfangswerte aufgeführt werden.
- Der Skalenparameter und, sofern die Antwort eine negative Binomialverteilung aufweist, der negative binomiale Parameter, müssen als letzte Anfangswerte angegeben werden.
- Wenn "Datei aufteilen" wirksam ist, müssen die Variablen mit der bzw. den Dateiteilungsvariablen beginnen (in der beim Erstellen der aufgeteilten Datei angegebenen Reihenfolge). Anschließend folgen *RowType\_*, *VarName\_*, *P1*, *P2* usw. wie oben angegeben. Aufteilungen müssen im angegebenen Dataset in derselben Reihenfolge erfolgen wie im ursprünglichen Dataset. .

*Hinweis:* Die Variablennamen  $P1$ ,  $P2$  usw. sind nicht erforderlich. Die Prozedur akzeptiert alle gültigen Variablennamen für die Parameter, da die Zuordnung von Variablen zu Parametern anhand der Position der Variablen und nicht anhand ihres Namens erfolgt. Alle Variablen nach dem letzten Parameter werden ignoriert.

Die Dateistruktur für die Anfangswerte stimmt mit der überein, die beim Exportieren des Modells als Daten verwendet wurde; daher können Sie die endgültigen Werte aus einer Ausführung der Prozedur als Eingabe für eine spätere Ausführung verwenden.

## Verallgemeinerte lineare Modelle: Statistiken

**Modelleffekte.** Die folgenden Optionen stehen zur Verfügung:

- **Analysetyp.** Geben Sie den Typ der zu erstellenden Analyse an. Eine Analyse des Typs I ist im Allgemeinen dann angebracht, wenn Sie von vorneherein Gründe dafür haben, die Prädiktoren im Modell zu ordnen. Typ III dagegen ist allgemeiner anwendbar. Wald-Statistiken bzw. Likelihood-Quotienten-Statistiken werden anhand der Auswahl in der Gruppe "Chi-Quadrat-Statistik" berechnet.
- **Konfidenzintervalle.** Geben Sie ein Konfidenzniveau größer als 50 und kleiner als 100 an. Wald-Intervalle beruhen auf der Annahme, dass die Parameter eine asymptotische Normalverteilung aufweisen. Profile-Likelihood-Intervalle sind genauer, können jedoch sehr rechenaufwendig sein. Die Toleranzstufe für Profil-Likelihood-Intervalle ist das Kriterium, anhand dessen der iterative Algorithmus zur Intervallberechnung gestoppt wird.
- **Log-Likelihood-Funktion.** Legt das Anzeigeformat der Log-Likelihood-Funktion fest. Die vollständige Funktion enthält einen zusätzlichen Term, der hinsichtlich der Parameterschätzungen konstant ist. Er hat keine Auswirkungen auf die Parameterschätzung und wird bei einigen Softwareprodukten nicht angezeigt.

**Drucken.** Die folgenden Ausgaben sind verfügbar:

- **Zusammenfassung der Fallverarbeitung.** Zeigt die Anzahl und den Prozentsatz der Fälle an, die in die Analyse und in die Tabelle "Korrelierte Datenzusammenfassung" aufgenommen bzw. daraus ausgeschlossen werden.
- **Deskriptive Statistiken.** Zeigt eine deskriptive Statistik und Zusammenfassungsinformationen über die abhängige Variable, die Kovariaten und die Faktoren an.
- **Modellinformationen.** Zeigt den Namen des Datasets, die abhängige Variable bzw. die Ereignis- und Versuchsvariablen, die Offsetvariable, die Skalengewichtungsvariable, die Wahrscheinlichkeitsverteilung und die Verknüpfungsfunktion an.
- **Statistik für Anpassungsgüte.** Zeigt an: Abweichung und skalierte Abweichung, Pearson-Chi-Quadrat und skaliertes Pearson-Chi-Quadrat, Log-Likelihood, Akaike-Informationskriterium (AIC), AIC mit Korrektur für endliche Stichproben (AICC), Bayes-Informationskriterium (BIC) und konsistentes AIC (CAIC).
- **Modellzusammenfassungsverstatistik.** Zeigt Tests für die Anpassungsgüte des Modells an, darunter Likelihood-Quotienten-Statistiken für den Omnibus-Test für die Anpassungsgüte, sowie Statistiken für Kontraste des Typs I bzw. III für jeden Effekt.
- **Parameterschätzungen.** Zeigt Parameterschätzungen und entsprechende Teststatistiken und Konfidenzintervalle an. Wahlweise können Sie zusätzlich zu den rohen, unbearbeiteten Parameterschätzungen auch potenzierte Parameterschätzungen anzeigen.
- **Kovarianzmatrix für Parameterschätzungen.** Zeigt die Kovarianzmatrix für die geschätzten Parameter an.
- **Korrelationsmatrix für Parameterschätzungen.** Zeigt die Korrelationsmatrix für geschätzte Parameter an.
- **Kontrastkoeffizientenmatrizen (L-Matrizen).** Zeigt die Kontrastkoeffizienten für die Standardeffekte und für die geschätzten Randmittel an, sofern auf der Registerkarte "Geschätzte Randmittel" angefordert.
- **Allgemeine schätzbare Funktionen.** Zeigt die Matrizen für die Generierung der Kontrastkoeffizientenmatrizen (L-Matrizen) an.

- **Iterationsverlauf.** Zeigt den Iterationsverlauf für Parameterschätzungen und Log-Likelihood an und druckt die letzte Evaluierung des Gradientenvektors und der Hesse-Matrix. Die Tabelle des Iterationsverlaufs zeigt Parameterschätzungen für jede  $n$ -te Iteration an, beginnend mit der 0-ten Iteration (Anfangsschätzungen), wobei  $n$  für den Wert des Druckintervalls steht. Wenn der Iterationsverlauf angefordert wird, wird die letzte Iteration unabhängig von  $n$  stets angezeigt.
- **Lagrange-Multiplikator-Test für metrische Parameter oder negative binomiale Hilfsparameter.** Zeigt die Statistiken für den Lagrange-Multiplikator-Test an, die zur Bewertung der Gültigkeit eines Skalensparameters dienen, der mithilfe des Pearson-Chi-Quadrats berechnet wurde oder für den bei der Normalverteilung, Gammaverteilung, inversen Normalverteilung und Tweedie-Verteilung ein fester Wert festgelegt wurde. Bei der negativen Binomialverteilung wird hiermit der feste Hilfsparameter getestet.

## Verallgemeinerte lineare Modelle: Geschätzte Randmittel

Auf dieser Registerkarte können Sie die geschätzten Randmittel für die Ebenen der Faktoren und die Faktorinteraktionen anzeigen. Des Weiteren können Sie den geschätzten Gesamtmittelwert anzeigen lassen. Geschätzte Randmittel sind für ordinal-multinomiale Modelle nicht verfügbar.

**Faktoren und Interaktionen.** Diese Liste enthält Faktoren, die auf der Registerkarte "Prädiktoren" angegeben wurden, und Interaktionen zwischen Faktoren, die auf der Registerkarte "Modell" angegeben wurden. Kovariaten sind aus dieser Liste ausgeschlossen. Terme können direkt aus dieser Liste ausgewählt werden oder mithilfe der Schaltfläche **Nach \*** zu einem Interaktionsterm kombiniert werden.

**Mittelwerte anzeigen für.** Geschätzte Mittelwerte werden für die ausgewählten Faktoren und Interaktionen zwischen Faktoren berechnet. Der Kontrast bestimmt, wie die Hypothesentests zum Vergleich der geschätzten Mittel eingerichtet werden. Für einfache Kontraste wird eine Referenzkategorie oder eine Faktorebene benötigt, mit der die anderen Elemente verglichen werden können.

- **Paarweise.** Paarweise Vergleiche werden für Kombinationen über alle Stufen der angegebenen oder implizierten Faktoren berechnet. Dies ist der einzige verfügbare Kontrast für Interaktionen zwischen Faktoren.
- **Einfach.** Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt.
- **Abweichung.** Jede Faktorstufe wird mit dem Gesamtmittelwert verglichen. Abweichungskontraste sind nicht orthogonal.
- **Differenz.** Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. Diese Kontraste werden auch als umgekehrte Helmert-Kontraste bezeichnet.
- **Helmert.** Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.
- **Wiederholt.** Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.
- **Polynomial.** Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

**Metrisch.** Geschätzte Randmittel können für die Antwort berechnet werden (auf der Grundlage der ursprünglichen Skala der abhängigen Variablen) oder für den linearen Prädiktor (auf der Grundlage der durch die Verknüpfungsfunktion transformierten abhängigen Variablen).

**Anpassung für Mehrfachvergleiche.** Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- **Bonferroni.** Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.

- *Bonferroni sequenziell*. Hierbei handelt es sich um ein sequentielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist was die Ablehnung einzelner Hypothesen angeht, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- *Sidak*. Dieses Verfahren liefert engere Grenzen als der Bonferroni-Ansatz.
- *Sidak (sequenziell)*. Hierbei handelt es sich um ein sequentielles, schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen angeht, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.

## Verallgemeinerte lineare Modelle: Speichern

Die markierten Elemente werden unter dem angegebenen Namen gespeichert. Sie können auswählen, dass vorhandene Variablen mit demselben Namen wie die neuen Variablen überschrieben werden sollen, oder Namenskonflikte durch das Anhängen von Suffixen, die die neuen Variablennamen eindeutig machen, vermeiden.

### Vorhergesagter Wert für den Mittelwert der abhängigen Variablen

Speichert vom Modell vorhergesagte Werte für jeden Fall in der ursprünglichen Antwortmetrik. Bei einer binomialen Antwortverteilung und einer binären abhängigen Variablen speichert die Prozedur vorhergesagte Wahrscheinlichkeiten. Bei einer multinomialen Antwortverteilung wird die Elementbeschriftung **Kumulative vorhergesagte Wahrscheinlichkeit** verwendet und die Prozedur speichert die kumulative vorhergesagte Wahrscheinlichkeit für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

### Untergrenze des Konfidenzintervalls für den Mittelwert der abhängigen Variablen

Speichert die Untergrenze des Konfidenzintervalls für den Mittelwert der Antwort. Bei einer multinomialen Antwortverteilung wird die Elementbeschriftung **Untergrenze des Konfidenzintervalls für die kumulativ vorhergesagte Wahrscheinlichkeit** verwendet und die Prozedur speichert die Untergrenze für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

### Obergrenze des Konfidenzintervalls für den Mittelwert der abhängigen Variablen

Speichert die Obergrenze des Konfidenzintervalls für den Mittelwert der Antwort. Bei einer multinomialen Antwortverteilung wird die Elementbeschriftung **Obergrenze des Konfidenzintervalls für die kumulativ vorhergesagte Wahrscheinlichkeit** verwendet und die Prozedur speichert die Obergrenze für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

### Vorhergesagte Kategorie

Bei Modellen mit Binomialverteilung und binärer abhängiger Variabler bzw. mit Multinomialverteilung wird hiermit die vorhergesagte Kategorie für die einzelnen Fälle gespeichert. Diese Option ist nicht für andere Antwortverteilungen verfügbar.

### Vorhergesagter Wert für den linearen Prädiktor

Speichert vom Modell vorhergesagte Werte für jeden Fall in der Metrik des linearen Prädiktors (transformierte Antwort über die angegebene Verknüpfungsfunktion). Bei einer multinomialen Antwortverteilung speichert die Prozedur den vorhergesagten Wert für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

### Geschätzter Standardfehler des vorhergesagten Werts der linearen Prädiktoren

Bei einer multinomialen Antwortverteilung speichert die Prozedur den geschätzten Standardfehler für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

Die folgenden Elemente sind bei einer multinomialen Antwortverteilung nicht verfügbar.

### Cook-Distanz

Ein Maß dafür, wie sehr sich die Residuen aller Fälle ändern würden, wenn ein bestimmter Fall aus der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, dass der Ausschluss eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substantiell verändert.

### Hebelwert

Misst den Einfluss eines Punktes auf die Anpassung der Regression. Der zentrierte Wert für die Hebelwirkung bewegt sich zwischen 0 (kein Einfluss auf die Anpassung) und  $(N-1)/N$ .

### Residuen

Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.

### Pearson-Residuum

Die Quadratwurzel des Beitrags eines Falls zur Pearson-Chi-Quadrat-Statistik, mit dem Vorzeichen des Rohresiduums.

### Standardisiertes Pearson-Residuum

Das Pearson-Residuum multipliziert mit der Quadratwurzel der Inversen des Produkts des Skalenparameters und 1 Hebelwert für den Fall.

### Abweichungsresiduum

Die Quadratwurzel des Beitrags eines Falls zur Abweichungsstatistik, mit dem Vorzeichen des Rohresiduums.

### Standardisiertes Abweichungsresiduum

Das Abweichungsresiduum multipliziert mit der Quadratwurzel der Inversen des Produkts des Skalenparameters und 1 Hebelwert für den Fall.

### Likelihood-Residuum

Die Quadratwurzel eines gewichteten Durchschnitts (auf der Grundlage des Hebelwert des Falls) der Quadrate der standardisierten Pearson- und Abweichungsresiduen mit dem Vorzeichen des Rohresiduums.

## Verallgemeinerte lineare Modelle: Exportieren

**Modell als Daten exportieren.** Schreibt ein Dataset im IBM SPSS Statistics-Format, das die Parameterkorrelations- oder -kovarianzmatrix enthält (mit Parameterschätzungen, Standardfehlern, Signifikanzwerten und Freiheitsgraden). Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **Teilungsvariablen.** Sofern verwendet, alle Variablen, die Aufteilungen (Splits) definieren.
- **RowType\_.** Nimmt Werte (und Wertbeschriftungen) *COV* (Kovarianzen), *CORR* (Korrelationen), *EST* (Parameterschätzungen), *SE* (Standardfehler), *SIG* (Signifikanzniveau) und *DF* (Freiheitsgrade der Stichprobenentnahme) an. Es gibt einen separaten Fall mit dem Zeilentyp *COV* (bzw. *CORR*) für jeden Modellparameter sowie einen separaten Fall für jeden der anderen Zeilentypen.
- **VarName\_.** Nimmt die Werte *P1*, *P2*, ..., die einer geordneten Liste aller geschätzten Modellparameter (mit Ausnahme der Parameter für den Maßstab oder der negativen Binomialparameter) entsprechen, für die Zeilentypen *COV* oder *CORR* an, wobei die Wertbeschriftungen den Parameterzeichenfolgen entsprechen, die in der Tabelle "Parameterschätzungen" angezeigt werden. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter (einschließlich der Maßstabswerte und der negativen binomialen Parameter) mit Variablenbeschriftungen, die den in der Tabelle "Parameterschätzungen" angezeigten Parameterzeichenfolgen entsprechen, und Werte entsprechend dem Zeilentyp.

Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzungen sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Für den Skalenparameter sind Kovarianzen, Korrelationen, Signifikanzniveau und Freiheitsgrade auf den systemdefiniert fehlenden Wert gesetzt. Wenn der Skalenparameter über Maximum Likelihood geschätzt wird, wird der Standardfehler ausgegeben, andernfalls wird der systemdefiniert fehlende Wert verwendet.

Für den negativen binomialen Parameter sind Kovarianzen, Korrelationen, Signifikanzniveau und Freiheitsgrade auf den systemdefiniert fehlenden Wert gesetzt. Wenn der negative binomiale Parameter über Maximum Likelihood geschätzt wird, wird der Standardfehler ausgegeben, andernfalls wird der systemdefiniert fehlende Wert verwendet.

Wenn Aufteilungen vorliegen, muss die Liste der Parameter über alle Splits akkumuliert sein. In einer Aufteilung können einige Parameter irrelevant sein; dies ist jedoch nicht mit redundant zu verwechseln. Für irrelevante Parameter sind alle Kovarianzen bzw. Korrelationen, Parameterschätzungen, Standardfehler, Signifikanzniveau und Freiheitsgrade auf den systemdefiniert fehlenden Wert gesetzt.

Sie können diese Matrixdatei als Anfangswerte für die weitere Modellschätzung verwenden. Beachten Sie, dass diese Datei nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar ist, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen. Doch auch dann sollten Sie darauf achten, dass alle Parameter in dieser Matrixdatei für die Prozedur, die die Datei einliest, dieselbe Bedeutung haben.

**Modell als XML exportieren.** Speichert die Parameterschätzungen und gegebenenfalls die Parameterkovarianzmatrix (falls ausgewählt) im XML-Format (PMML). Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden. .

## Zusätzliche Funktionen beim Befehl GENLIN

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl CRITERIA können Sie Anfangswerte für Parameterschätzungen als Zahlenliste angeben.
- Mit dem Unterbefehl EMMEANS können Sie bei der Berechnung der geschätzten Randmittel einen anderen Wert für die Kovariaten festlegen als den Mittelwert.
- Mit dem Unterbefehl EMMEANS können Sie außerdem benutzerdefinierte polynomiale Kontraste für geschätzte Randmittel angeben.
- Mit den Schlüsselwörtern TABLES und COMPARE des Unterbefehls EMMEANS können Sie ein Subset der Faktoren, für die geschätzte Randmittel angezeigt werden, mithilfe des angegebenen Kontrasttyps vergleichen lassen.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Verallgemeinerte Schätzungsgleichungen

Die Prozedur "Verallgemeinerte Schätzungsgleichungen" erweitert das verallgemeinerte lineare Modell zur Berücksichtigung der Analyse wiederholter Messungen oder anderer korrelierter Beobachtungen, wie beispielsweise in Clustern gruppierter Daten.

**Beispiel.** Öffentliche Gesundheitsbeamte können generalisierte Schätzungsgleichungen verwenden, um eine wiederholte Messung der logistischen Regression einzufügen, um Auswirkungen der Luftverschmutzung auf Kinder zu untersuchen.

Erläuterungen der Daten für verallgemeinerte Schätzungsgleichungen

**Daten.** Die Antwort kann metrisch (Skala), eine Anzahl, binär oder vom Typ "Ereignisse in Versuchen" sein. Es wird davon ausgegangen, dass Faktoren kategorial sind. Es wird davon ausgegangen, dass Kovariaten, Skalengewichtung und Offset metrisch (Skala) sind. Variablen, die zur Definition von Subjekten oder von wiederholten Messungen innerhalb von Subjekten verwendet werden, können nicht verwendet werden, um die Antwort zu definieren, sie können jedoch andere Funktionen im Modell wahrnehmen.

**Annahmen.** Es wird davon ausgegangen, dass die Fälle innerhalb der Subjekte abhängig und zwischen den verschiedenen Objekten unabhängig sind. Die Korrelationsmatrix, die die Innersubjektabhängigkeiten darstellt, wird als Teil des Modells geschätzt.

Ermitteln von verallgemeinerten Schätzungsgleichungen

Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Verallgemeinerte lineare Modelle > Verallgemeinerte Schätzungsgleichungen...**

1. Wählen Sie mindestens eine Subjektvariable aus (Informationen zu weiteren Optionen finden Sie weiter unten).

Die Wertekombination der angegebenen Variablen sollte die **Subjekte** innerhalb des Datensets eindeutig definieren. Beispiel: Eine einzelne Variable *Patienten-ID* sollte ausreichen, um die Subjekte in einem einzelnen Krankenhaus zu definieren, doch die Kombination aus *Krankenhaus-ID* und *Patienten-ID* kann erforderlich sein, wenn die Patienten-IDs nicht krankenhausesübergreifend eindeutig sind. Bei einer Einstellung mit wiederholten Messungen werden für jedes Subjekt mehrere Beobachtungen aufgezeichnet, sodass jedes Subjekt mehrere Fälle im Dataset belegen kann.

2. Geben Sie auf der Registerkarte Typ des Modells eine Verteilungs- und Verknüpfungsfunktion an.
3. Wählen Sie auf der Registerkarte Abhängige Variable eine abhängige Variable aus.
4. Wählen Sie auf der Registerkarte Prädiktoren die Faktoren und Kovariaten aus, die zur Vorhersage der abhängigen Variablen verwendet werden sollen.
5. Geben Sie auf der Registerkarte Modell mithilfe der ausgewählten Faktoren und Kovariaten Modelleffekte an.

Wahlweise können Sie auf der Registerkarte "Wiederholt" folgende Elemente angeben:

**Innersubjektvariablen.** Die Kombination der Werte von Innersubjektvariablen legt die Reihenfolge der Messungen innerhalb von Subjekten fest; dadurch legt die Kombination von Innersubjekt- und Subjektvariablen jede Messung eindeutig fest. So definiert beispielsweise die Kombination aus *Zeitraum*, *Krankenhaus-ID* und *Patienten-ID* für jeden Fall einen ganz bestimmten Aufenthalt eines bestimmten Patienten in einem bestimmten Krankenhaus.

Wenn das Dataset bereits so sortiert ist, dass die wiederholten Messungen für jedes Subjekt in einem zusammenhängenden Block von Fällen und in der richtigen Reihenfolge vorliegen, ist es streng genommen nicht erforderlich, eine Innersubjektvariable anzugeben und Sie können die Auswahl der Option **Fälle nach Subjekt- und Innersubjektvariablen sortieren** aufheben und die für die Durchführung der (temporären) Sortierung erforderliche Prozessorzeit einsparen. Im Allgemeinen ist es sinnvoll, die Innersubjektvariablen zu verwenden, um die richtige Sortierung der Messungen sicherzustellen.

Subjekt- und Innersubjektvariablen können nicht für die Definition der Antwort verwendet werden, sie können jedoch andere Funktionen im Modell wahrnehmen. So könnte *Krankenhaus-ID* beispielsweise als Faktor im Modell verwendet werden.

**Kovarianzmatrix.** Der modellbasierte Schätzer ist das Negative der verallgemeinerten Inversen der Hesse-Matrix. Der robuste Schätzer (auch Huber-/White-/Sandwich-Schätzer genannt) ist ein "korrigierter" modellbasierter Schätzer, der eine konsistente Schätzung der Kovarianz bietet, selbst wenn die Arbeitskorrelationsmatrix falsch angegeben wurde. Diese Spezifikation gilt für die Parameter im Bereich des linearen Modells der verallgemeinerten Schätzungsgleichungen, während die Spezifikation auf der Registerkarte Schätzung nur für das ursprüngliche verallgemeinerte lineare Modell gilt.

**Arbeitskorrelationsmatrix.** Diese Korrelationsmatrix stellt die Innersubjektabhängigkeiten dar. Ihre Größe wird durch die Anzahl der Messungen und damit durch die Kombination der Werte von Innersubjektvariablen bestimmt. Sie können eine der folgenden Strukturen angeben:

- **Unabhängig.** Wiederholte Messungen sind unkorreliert.
- **AR(1).** Wiederholte Messungen weisen eine autoregressive Beziehung erster Ordnung auf. Die Korrelation zwischen den beiden Elementen entspricht  $\rho$  für benachbarte Elemente,  $\rho^2$  für Elemente, die durch ein Drittel getrennt sind, ist so eingeschränkt, dass  $-1 < \rho < 1$ .
- **Austauschbar.** Diese Struktur weist homogene Korrelationen zwischen Elementen auf. Diese Struktur ist auch als zusammengesetzt symmetrische Struktur bekannt.
- **M-abhängig.** Aufeinander folgende Messungen haben einen gemeinsamen Korrelationskoeffizienten, Messungspaare, die durch eine dritte Messung getrennt sind, haben einen gemeinsamen Korrelationskoeffizienten usw., bis hin zu Messungspaaren, die durch  $m-1$  andere Messungen getrennt sind. Angenommen, Schüler von der 3. bis zur 7. Klasse machen jedes Jahr standardisierte Prüfungen. Diese Struktur setzt voraus, dass die Notenscores für die 3. und 4., die 4. und 5., die 5. und 6. und die 6. und 7. Klasse dieselbe Korrelation haben. Für die 3. und 5., 4. und 6. und 5. und 7. Klasse gilt dieselbe Korrelation; für die 3. und 6. und die 4. und 7. Klasse gilt dieselbe Korrelation. Messungen mit einem Abstand größer als  $m$  gelten als unkorreliert. Geben Sie bei der Auswahl dieser Struktur einen Wert für  $m$  an, der niedriger ist als die Ordnung der Arbeitskorrelationsmatrix.

- **Unstrukturiert.** Eine ganz allgemeine Korrelationsmatrix.

Standardmäßig passt diese Prozedur die Korrelationsschätzungen um die Anzahl der nicht redundanten Parameter an. Der Verzicht auf diese Anpassung kann sinnvoll sein, wenn die Schätzungen gegenüber Replikationsänderungen der Daten auf der Subjektebene invariant sein sollen.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus für verallgemeinerte Schätzungsgleichungen vorgenommen werden. Geben Sie eine nicht negative Ganzzahl an. Diese Spezifikation gilt für die Parameter im Bereich des linearen Modells der verallgemeinerten Schätzungsgleichungen, während die Spezifikation auf der Registerkarte Schätzung nur für das ursprüngliche verallgemeinerte lineare Modell gilt.
- **Matrix aktualisieren.** Die Elemente in der Arbeitskorrelationsmatrix werden anhand der Parameterschätzungen geschätzt, die bei jeder Iteration des Algorithmus aktualisiert werden. Wenn die Arbeitskorrelationsmatrix überhaupt nicht aktualisiert wird, wird die ursprüngliche Arbeitskorrelationsmatrix während des gesamten Schätzungsvorgangs verwendet. Falls die Matrix aktualisiert wird, können Sie das Iterationsintervall angeben, in dem die Elemente der Arbeitskorrelationsmatrix aktualisiert werden sollen. Durch die Angabe eines Werts größer 1 kann eventuell die Verarbeitungszeit reduziert werden.

**Konvergenzkriterien.** Diese Spezifikationen gelten für die Parameter im Bereich des linearen Modells der verallgemeinerten Schätzungsgleichungen, während die Spezifikation auf der Registerkarte Schätzung nur für das ursprüngliche verallgemeinerte lineare Modell gilt.

- **Parameterkonvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzungen unter dem angegebenen (positiven) Wert liegt.
- **Konvergenz der Hesse-Matrix.** Konvergenz wird angenommen, wenn eine auf der Hesse-Matrix beruhende Statistik unterhalb des angegebenen Werts (dieser muss positiv sein) liegt.

## Verallgemeinerte Schätzungsgleichungen: Typ des Modells

Auf der Registerkarte "Typ des Modells" können Sie die Verteilung und die Verknüpfungsfunktion für Ihr Modell angeben, die Schnellverfahren für mehrere häufig verwendete Modelle bieten, die durch den Antworttyp kategorisiert sind.

Modelltypen

**Metrische abhängige Variable.** Die folgenden Optionen stehen zur Verfügung:

- **Linear.** Gibt "Normal" als Verteilung und "Identität" als Verknüpfungsfunktion an.
- **Gamma mit Log-Verknüpfung.** Gibt "Gamma" als Verteilung und "Log" als Verknüpfungsfunktion an.

**Ordinale abhängige Variable.** Die folgenden Optionen stehen zur Verfügung:

- **Ordinal logistisch.** Gibt "Multinomial (ordinal)" als Verteilung und "Logit (kumulativ)" als Verknüpfungsfunktion an.
- **Ordinal Probit.** Gibt "Multinomial (ordinal)" als Verteilung und "Probit (kumulativ)" als Verknüpfungsfunktion an.

**Häufigkeiten.** Die folgenden Optionen stehen zur Verfügung:

- **Poisson loglinear.** Gibt "Poisson" als Verteilung und "Log" als Verknüpfungsfunktion an.
- **Negativ binomial mit Log-Verknüpfung.** Gibt "Negativ binomial" (mit einem Wert von 1 für den Hilfsparameter) als Verteilung und "Log" als Verknüpfungsfunktion an. Damit die Prozedur den Wert des Hilfsparameters schätzt, geben Sie ein benutzerdefiniertes Modell mit negativer binomialer Verteilung an und wählen Sie **Wert schätzen** in der Gruppe "Parameter".

**Binär oder Anzahl Ereignisse.** Die folgenden Optionen stehen zur Verfügung:

- **Binär logistisch.** Gibt "Binomial" als Verteilung und "Logit" als Verknüpfungsfunktion an.
- **Binär Probit.** Gibt "Binomial" als Verteilung und "Probit" als Verknüpfungsfunktion an.
- **Intervallzensiertes Überleben.** Gibt "Binomial" als Verteilung und "Log-Log komplementär" als Verknüpfungsfunktion an.

**Mischung.** Die folgenden Optionen stehen zur Verfügung:

- **Tweedie mit Verknüpfungsfunktion "Log".** Gibt "Tweedie" als Verteilung und "Log" als Verknüpfungsfunktion an.
- **Tweedie mit Verknüpfungsfunktion "Identität".** Gibt "Tweedie" als Verteilung und "Identität" als Verknüpfungsfunktion an.

**Benutzerdefiniert.** Geben Sie Ihre eigene Kombination von Verteilung und Verknüpfungsfunktion an.

Verteilung

Diese Auswahl gibt die Verteilung der abhängigen Variablen an. Die Möglichkeit einer anderen Verteilung als "Normal" und einer anderen Verknüpfungsfunktion als "Identität" ist die wichtigste Verbesserung des verallgemeinerten linearen Modells gegenüber dem allgemeinen linearen Modell. Es gibt zahlreiche mögliche Kombinationen aus Verteilung und Verknüpfungsfunktion und es können mehrere davon für das jeweils vorliegende Dataset geeignet sein. Daher können Sie sich in Ihrer Wahl durch theoretische Vorüberlegungen leiten lassen oder davon, welche Kombination am besten zu passen scheint.

- **Binomial.** Diese Verteilung ist nur für Variablen geeignet, die eine binäre Antwort oder eine Anzahl von Ereignissen repräsentieren.
- **Gamma.** Diese Verteilung eignet sich für Variablen mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Wenn ein Datenwert kleiner oder gleich 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Invers normal.** Diese Verteilung eignet sich für Variablen mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Wenn ein Datenwert kleiner oder gleich 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Negativ binomial.** Diese Verteilung lässt sich als Anzahl der Versuche betrachten, die erforderlich sind, um  $k$  Erfolge zu beobachten, und eignet sich für Variablen mit nicht negativen ganzzahligen Werten. Wenn ein Datenwert keine Ganzzahl oder kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet. Der Wert für den Hilfsparameter der negativen Binomialverteilung kann eine beliebige Zahl größer als oder gleich 0 sein. Sie können ihn auf einen festen Wert setzen oder ihn nach dem Verfahren schätzen lassen. Wenn der Hilfsparameter auf 0 gesetzt wird, ist die Verwendung dieser Verteilung äquivalent zur Verwendung der Poisson-Verteilung.
- **Normal.** Diese Option eignet sich für metrische Variablen, deren Werte eine symmetrische, glockenförmige Verteilung um einen Mittelwert aufweisen. Die abhängige Variable muss numerisch sein.
- **Poisson-Verteilung.** Diese Verteilung lässt sich als Anzahl der Vorkommen eines untersuchten Ereignisses in einem festen Zeitraum betrachten und eignet sich für Variablen mit nicht negativen ganzzahligen Werten. Wenn ein Datenwert keine Ganzzahl oder kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Tweedie.** Diese Verteilung eignet sich für Variablen, die durch Poisson-Mischungen von Gammaverteilungen repräsentiert werden können; die Verteilung ist dahingehend "gemischt", dass sie sowohl Eigenschaften von stetigen Verteilungen (nimmt nicht negative reelle Werte an) als auch von diskreten Verteilungen (positive Wahrscheinlichkeitsmasse an einem Einzelwert, 0) aufweist. Die abhängige Variable muss numerisch sein, mit Datenwerten größer oder gleich 0. Wenn ein Datenwert kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet. Der feste Wert des Parameters der Tweedie-Verteilung kann jede beliebige Zahl zwischen 1 und 2 sein.
- **Multinomial.** Diese Verteilung eignet sich für Variablen, die eine ordinale abhängige Variable repräsentieren. Bei der abhängigen Variablen kann es sich um eine numerische Variable oder eine Zeichenfolgevariable handeln; sie muss mindestens zwei verschiedene gültige Datenwerte aufweisen.

Verknüpfungsfunktion

Die Verknüpfungsfunktion ist eine Transformation der abhängigen Variable, die eine Schätzung des Modells ermöglicht. Die folgenden Funktionen sind verfügbar:

- **Identität.**  $f(x)=x$ . Die abhängige Variable ist nicht transformiert. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Komplementäres Protokoll-Protokoll.**  $f(x)=\log(-\log(1-x))$ . Nur für die Binomialverteilung geeignet.

- **Kumulative Cauchit.**  $f(x) = \tan(\pi(x - 0,5))$ , angewendet auf die kumulative Wahrscheinlichkeit für jede Kategorie der Antwort. Nur für die Multinomialverteilung geeignet.
- **Kumulatives komplementäres Protokoll-Protokoll.**  $f(x) = \ln(-\ln(1-x))$ , angewendet auf die kumulative Wahrscheinlichkeit für jede Kategorie der Antwort. Nur für die Multinomialverteilung geeignet.
- **Logit (kumulativ).**  $f(x) = \ln(x / (1-x))$ , auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Nur für die Multinomialverteilung geeignet.
- **Kumulatives negatives Protokoll-Protokoll.**  $f(x) = -\ln(-\ln(x))$ , angewendet auf die kumulative Wahrscheinlichkeit für jede Kategorie der Antwort. Nur für die Multinomialverteilung geeignet.
- **Probit (kumulativ).**  $f(x) = \Phi^{-1}(x)$ , auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Dabei steht  $\Phi^{-1}$  für die umgekehrte kumulative Standardnormalverteilungsfunktion steht. Nur für die Multinomialverteilung geeignet.
- **Log.**  $f(x) = \log(x)$ . Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Protokollergänz.**  $f(x) = \log(1-x)$ . Nur für die Binomialverteilung geeignet.
- **Logit.**  $f(x) = \log(x/(1-x))$ . Nur für die Binomialverteilung geeignet.
- **Negatives Binomial.**  $f(x) = \log(x / (x+k^{-1}))$ . Hierbei steht  $k$  für den Nebenparameter der negativen Binomialverteilung. Nur für die negative Binomialverteilung geeignet.
- **Negatives Protokoll-Protokoll.**  $f(x) = -\log(-\log(x))$ . Nur für die Binomialverteilung geeignet.
- **-Odds Power.**  $f(x) = [(x/(1-x))^\alpha - 1] / \alpha$ , wenn  $\alpha \neq 0$  ist.  $f(x) = \log(x)$ , wenn  $\alpha = 0$  ist.  $\alpha$  ist die erforderliche Zahlenangabe und muss eine reelle Zahl sein. Nur für die Binomialverteilung geeignet.
- **Probit.**  $f(x) = \Phi^{-1}(x)$ , wobei  $\Phi^{-1}$  die umgekehrte kumulative Standardnormalverteilungsfunktion ist. Nur für die Binomialverteilung geeignet.
- **Power.**  $f(x) = x^\alpha$ , wenn  $\alpha \neq 0$  ist.  $f(x) = \log(x)$ , wenn  $\alpha = 0$  ist.  $\alpha$  ist die erforderliche Zahlenangabe und muss eine reelle Zahl sein. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.

## Verallgemeinerte Schätzungsgleichungen: Antwort

In vielen Fällen können Sie einfach eine abhängige Variable angeben. Allerdings erfordern Variablen, die nur zwei Werte annehmen, und Antworten, die Ereignisse in Versuchen aufzeichnen, besondere Aufmerksamkeit.

- **Binäre Antwort.** Wenn die abhängige Variable nur zwei Werte annimmt, können Sie die Referenzkategorie für die Parameterschätzung angeben. Binäre Antwortvariablen können Zeichenfolgevariablen oder numerische Variablen sein.
- **Anzahl der Ereignisse, die in einem Set von Versuchen vorkommen.** Wenn die Antwort eine Anzahl von Ereignissen ist, die in einem Set von Versuchen vorkommen, enthält die abhängige Variable die Anzahl der Ereignisse und Sie können eine zusätzliche Variable auswählen, die die Anzahl der Versuche enthält. Wenn die Anzahl der Versuche über alle Subjekte gleich ist, können die Versuche alternativ auch über einen festen Wert angegeben werden. Die Anzahl der Versuche sollte größer oder gleich der Anzahl der Ereignisse für jeden Fall sein. Bei den Ereignissen sollte es sich um nicht negative Ganzzahlen und bei den Versuchen um positive Ganzzahlen handeln.

Bei ordinal-multinomialen Modellen können Sie die Kategoriereihenfolge der Antwort angeben: aufsteigend, absteigend bzw. Daten (die Reihenfolge "Daten" bedeutet, dass der erste in den Daten gefundene Wert die erste Kategorie definiert und der letzte gefundene Wert die letzte Kategorie).

**Skalengewichtung.** Der Skalenparameter ist ein geschätzter Modellparameter, der mit der Varianz der Antwort zusammenhängt. Die Skalengewichtungen sind "bekannte" Werte, die sich zwischen den einzelnen Beobachtungen unterscheiden können. Wenn die Skalengewichtungsvariable angegeben ist, wird der Skalenparameter, der mit der Varianz der Antwort zusammenhängt, für jede Beobachtung durch diese Variable geteilt. Fälle, deren Skalengewichtungswerte kleiner-gleich 0 sind oder fehlen, werden in der Analyse nicht verwendet.

## Verallgemeinerte Schätzungsgleichungen: Referenzkategorie

Bei binären Antworten können Sie die Referenzkategorie für die abhängige Variable auswählen. Dies kann sich auf bestimmte Ausgaben, wie beispielsweise Parameterschätzungen und gespeicherte Werte, auswirken, sollte jedoch nicht die Anpassungsgüte des Modells verändern. Beispiel: Angenommen, Ihre binäre Antwort nimmt die Werte 0 und 1 an:

- Standardmäßig verwendet die Prozedur die letzte Kategorie (die mit dem höchsten Wert), also 1, als Referenzkategorie. In dieser Situation wird anhand der im Modell gespeicherten Wahrscheinlichkeitswerte die Wahrscheinlichkeit geschätzt, dass ein bestimmter Fall den Wert 0 annimmt, und die Parameterschätzungen sollten als in Bezug zur Likelihood von Kategorie 0 betrachtet werden.
- Wenn Sie die erste Kategorie (die mit dem niedrigsten Wert), also 0, als Referenzkategorie angeben, wird anhand der im Modell gespeicherten Wahrscheinlichkeitswerte die Wahrscheinlichkeit geschätzt, dass ein bestimmter Fall den Wert 1 annimmt.
- Wenn Sie die benutzerdefinierte Kategorie angeben und für Ihre Variable Beschriftungen definiert sind, können Sie die Referenzkategorie durch Auswahl eines Werts aus der Liste festlegen. Dies kann nützlich sein, wenn Sie bei der Festlegung eines Modells nicht mehr wissen, wie genau eine bestimmte Variable codiert war.

## Verallgemeinerte Schätzungsgleichungen: Prädiktoren

Auf der Registerkarte "Prädiktoren" können Sie die Faktoren und Kovariaten angeben, die zum Erstellen der Modelleffekte und zur Angabe eines optionalen Offsets verwendet werden.

**Faktoren.** Faktoren sind kategoriale Prädiktoren; es kann sich dabei um numerische Variablen oder Zeichenfolgevariablen handeln.

**Kovariaten.** Kovariaten sind metrische Prädiktoren; sie müssen numerisch sein.

*Hinweis:* Wenn die Antwort binomial mit Binärformat ist, berechnet die Prozedur Abweichungs- und Chi-Quadrat-Anpassungsstatistiken anhand von Teilgesamtheiten, die auf der Kreuzklassifikation beobachteter Werte der ausgewählten Faktoren und Kovariaten beruhen. Sie sollten dasselbe Set an Prädiktoren während mehrerer Ausführungen der Prozedur beibehalten, um eine konsistente Anzahl an Teilgesamtheiten zu gewährleisten.

**Offset.** Der Term "Offset" ist ein "struktureller" Prädiktor. Sein Koeffizient wird nicht vom Modell geschätzt, sondern es wird davon ausgegangen, dass er den Wert 1 aufweist. Daher werden die Werte des Offsets einfach zum linearen Prädiktor des Ziels addiert. Dies ist besonders nützlich bei Poisson-Regressionsmodellen, bei denen die verschiedenen Fälle dem relevanten Ereignis unterschiedlich stark ausgesetzt sein können.

Beispielsweise gibt es bei der Modellierung der Unfallraten für einzelne Fahrer einen wichtigen Unterschied zwischen einem Fahrer, der in 3 Jahren Fahrpraxis einen Unfall verursacht hat und einem Fahrer, der in 25 Jahren einen Unfall verursacht hat. Die Anzahl der Unfälle kann als Poisson- oder negative Binomialantwort mit einer Protokollverknüpfung modelliert werden, wenn der natürliche Logarithmus der Fahrpraxis des Fahrers als Offset-Term eingeschlossen wird.

Andere Kombinationen der Verteilung und Verknüpfungstypen machen andere Transformationen der Offsetvariable erforderlich.

## Verallgemeinerte Schätzungsgleichungen: Optionen

Diese Optionen werden auf alle auf der Registerkarte "Prädiktoren" angegebenen Faktoren angewendet.

**Benutzerdefiniert fehlende Werte.** Faktoren müssen gültige Werte für einen Fall aufweisen, um in die Analyse aufgenommen zu werden. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Faktorvariablen als gültige Werte behandelt werden sollen.

**Kategoriereihenfolge.** Diese ist relevant für die Bestimmung der letzten Stufe eines Faktors, die einem redundanten Parameter im Schätzalgorithmus zugeordnet werden kann. Durch das Ändern der Kategoriereihenfolge können sich die Werte der Faktorstufeneffekte ändern, da diese Parameterschätzungen bezogen auf die letzte Stufe berechnet werden. Die Faktoren können in aufsteigender Reihenfolge vom

niedrigsten zum höchsten Wert, in absteigender Reihenfolge vom höchsten zum niedrigsten Wert oder in "Datenreihenfolge" angeordnet werden. Dies bedeutet, dass der erste Wert, der in den Daten vorgefunden wird, die erste Kategorie definiert und der letzte eindeutige Wert, der vorgefunden wird, die letzte Kategorie.

## Verallgemeinerte Schätzungsgleichungen: Modell

**Modelleffekte angeben.** Das Standardmodell ist "Nur konstanter Term". Daher müssen Sie die anderen Modelleffekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme erstellen.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

**Haupteffekte.** Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

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

**Mehrfaktoriell.** Erstellt alle möglichen Interaktionen und Haupteffekte der ausgewählten Variablen.

**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.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell erstellen. Verschachtelte Terme sind nützlich, um den Effekt von Faktoren oder Kovariaten zu analysieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. Eine Lebensmittelkette kann beispielsweise das Kaufverhalten ihrer Kunden in mehreren Filialen untersuchen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

Darüber hinaus können Sie Interaktionseffekte einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

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

- Alle Faktoren innerhalb einer Interaktion müssen eindeutig sein. Dementsprechend ist die Angabe von  $A*A$  unzulässig, wenn  $A$  ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von  $A(A)$  unzulässig, wenn  $A$  ein Faktor ist.
- Effekte dürfen nicht in einer Kovariaten verschachtelt werden. Dementsprechend ist die Angabe von  $A(X)$  unzulässig, wenn  $A$  ein Faktor und  $X$  eine Kovariante ist.

**Konstanter Term.** Der konstante Term wird gewöhnlich in das Modell aufgenommen. Wenn anzunehmen ist, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

Modelle mit multinomialer ordinaler Verteilung weisen nicht einen einzelnen konstanten Term auf, sondern es gibt dort Schwellenwertparameter, die Übergangspunkte zwischen benachbarten Kategorien definieren. Die Schwellenwerte werden immer in das Modell aufgenommen.

## Verallgemeinerte Schätzungsgleichungen: Schätzung

**Parameterschätzung.** Mit den Steuerelementen in dieser Gruppe können Sie Schätzmethoden festlegen und Anfangswerte für die Parameterschätzungen angeben.

- **Methode.** Sie können eine Methode für die Parameterschätzung auswählen. Sie haben die Wahl zwischen "Newton-Raphson", "Fisher-Scoring" und einer Hybridmethode, bei der zuerst Iterationen des Fisher-Scoring durchgeführt werden und dann zur Methode "Newton-Raphson" gewechselt wird. Wenn

während der Phase "Fisher-Scoring" der Hybridmethode Konvergenz erreicht wird, bevor die maximale Anzahl an Fisher-Iterationen erreicht wurde, fährt der Algorithmus mit der Newton-Raphson-Methode fort.

- **Skalenparametermethode.** Sie können eine Schätzmethode für den Skalenparameter auswählen.

Bei der Maximum-Likelihood-Methode wird der Skalenparameter zusammen mit den Modelleffekten geschätzt; beachten Sie, dass diese Option ungültig ist, wenn die Antwort eine negative Binomialverteilung, eine Poisson-Verteilung oder eine Binomialverteilung aufweist. Da das Konzept der Likelihood in verallgemeinerten Schätzungsgleichungen nicht vorkommt, gilt diese Spezifikation nur für das ursprüngliche verallgemeinerte lineare Modell. Die betreffende Skalenparameterschätzung wird dann an die verallgemeinerten Schätzungsgleichungen weitergeleitet, die den Skalenparameter anhand des Pearson-Chi-Quadrats, dividiert durch die Anzahl der Freiheitsgrade, aktualisieren.

Die Optionen "Abweichung" und "Pearson-Chi-Quadrat" schätzen den Skalenparameter aus dem Wert dieser Statistiken im ursprünglichen verallgemeinerten linearen Modell. Diese Skalenparameterschätzung wird dann an die verallgemeinerten Schätzungsgleichungen weitergeleitet, die ihn als fest behandeln.

Alternativ können Sie einen festen Wert für den Skalenparameter angeben. Dieser wird beim Schätzen des ursprünglichen verallgemeinerten linearen Modells und in den verallgemeinerten Schätzungsgleichungen als fest behandelt.

- **Anfangswerte.** Die Prozedur berechnet automatisch Anfangswerte für Parameter. Alternativ können Sie Anfangswerte für die Parameterschätzungen angeben.

Die auf dieser Registerkarte angegebenen Iterationen und Konvergenzkriterien können nur auf das ursprüngliche verallgemeinerte lineare Modell angewendet werden. Informationen zu den bei der Anpassung verallgemeinerter Schätzungsgleichungen verwendeten Schätzkriterien finden Sie auf der Registerkarte Wiederholt.

**Iterationen.** Die folgenden Optionen stehen zur Verfügung:

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nicht negative Ganzzahl an.
- **Maximalzahl für Schritthalbierung.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritthalbierung erreicht ist. Geben Sie eine positive Ganzzahl an.
- **Prüfen auf Datenpunkttrennung.** Mit dieser Option lassen Sie Tests durch den Algorithmus durchführen, mit denen sichergestellt wird, dass die Parameterschätzungen eindeutige Werte aufweisen. Eine Trennung wird vorgenommen, sobald ein Modell erzeugt werden kann, in dem alle Fälle fehlerfrei klassifiziert werden. Diese Option ist für multinomiale Antworten und binomiale Antworten mit Binärformat verfügbar.

**Konvergenzkriterien.** Die folgenden Optionen sind verfügbar:

- **Parameterkonvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzungen unter dem angegebenen (positiven) Wert liegt.
- **Log-Likelihood-Konvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei der Log-Likelihood-Funktion unter dem angegebenen (positiven) Wert liegt.
- **Konvergenz der Hesse-Matrix.** Für die Spezifikation 'Absolut' wird angenommen, dass eine Konvergenz vorliegt, wenn eine Statistik auf der Basis der Konvergenz der Hesse-Matrix kleiner als der angegebene positive Wert ist. Für die Spezifikation "Relativ" wird angenommen, dass eine Konvergenz vorliegt, wenn die Statistik kleiner als das Produkt aus dem angegebenen positiven Wert und dem absoluten Wert der Log-Likelihood ist.

**Toleranz für Prüfung auf Singularität.** Singuläre (bzw. nicht invertierbare) Matrizen weisen linear abhängige Spalten auf, die zu ernststen Problemen für den Schätzalgorithmus führen können. Auch annähernd singuläre Matrizen können zu schlechten Ergebnissen führen, daher behandelt die Prozedur eine Matrix, deren Determinante unter dem Toleranzwert liegt, als singulär. Geben Sie einen positiven Wert an.

## Verallgemeinerte Schätzungsgleichungen: Anfangswerte

Die Prozedur schätzt ein ursprüngliches verallgemeinertes lineares Modell. Die Schätzungen aus diesem Modell werden dann als Anfangswerte für die Parameterschätzungen im Bereich des linearen Modells der verallgemeinerten Schätzungsgleichungen verwendet. Anfangswerte sind für die Arbeitskorrelationsmatrix nicht erforderlich, da die Matrixelemente auf den Parameterschätzungen beruhen. Die in diesem Dialogfeld angegebenen Anfangswerte werden als Ausgangspunkt für das ursprüngliche verallgemeinerte lineare Modell verwendet und nicht für die verallgemeinerten Schätzungsgleichungen, es sei denn, der Wert für "Maximalzahl der Iterationen" auf der Registerkarte Schätzung ist auf 0 gesetzt.

Wenn Anfangswerte angegeben wurden, müssen diese für alle Parameter (auch redundante Parameter) im Modell angegeben werden. Im Dataset müssen die Variablen von links nach rechts folgende Reihenfolge aufweisen: *RowType\_*, *VarName\_*, *P1*, *P2* usw. Dabei sind *RowType\_* und *VarName\_* Zeichenfolgevariablen und *P1*, *P2* usw. numerische Variablen, die einer geordneten Liste der Parameter entsprechen.

- Die Anfangswerte werden in einem Datensatz angegeben, mit dem Wert *EST* für die Variable *RowType\_*. Die tatsächlichen Anfangswerte werden unter den Variablen *P1*, *P2* usw. angegeben. Die Prozedur ignoriert alle Datensätze, für die *RowType\_* einen anderen Wert als *EST* aufweist sowie alle Datensätze nach dem ersten Vorkommen von *RowType\_*, die gleich *EST* sind.
- Der konstante Term, sofern im Modell enthalten, bzw. die Schwellenwertparameter, falls das Modell eine Multinomialverteilung aufweist, müssen als erste Anfangswerte aufgeführt werden.
- Der Skalenparameter und, sofern die Antwort eine negative Binomialverteilung aufweist, der negative binomiale Parameter, müssen als letzte Anfangswerte angegeben werden.
- Wenn "Datei aufteilen" wirksam ist, müssen die Variablen mit der bzw. den Dateiteilungsvariablen beginnen (in der beim Erstellen der aufgeteilten Datei angegebenen Reihenfolge). Anschließend folgen *RowType\_*, *VarName\_*, *P1*, *P2* usw. wie oben angegeben. Aufteilungen müssen im angegebenen Dataset in derselben Reihenfolge erfolgen wie im ursprünglichen Dataset. .

*Hinweis:*: Die Variablennamen *P1*, *P2* usw. sind nicht erforderlich. Die Prozedur akzeptiert alle gültigen Variablennamen für die Parameter, da die Zuordnung von Variablen zu Parametern anhand der Position der Variablen und nicht anhand ihres Namens erfolgt. Alle Variablen nach dem letzten Parameter werden ignoriert.

Die Dateistruktur für die Anfangswerte stimmt mit der überein, die beim Exportieren des Modells als Daten verwendet wurde; daher können Sie die endgültigen Werte aus einer Ausführung der Prozedur als Eingabe für eine spätere Ausführung verwenden.

## Verallgemeinerte Schätzungsgleichungen: Statistiken

**Modelleffekte.** Die folgenden Optionen sind verfügbar:

- **Analysetyp.** Geben Sie den Typ der zum Testen der Modelleffekte zu erstellenden Analyse an. Eine Analyse des Typs I ist im Allgemeinen dann angebracht, wenn Sie von vorneherein Gründe dafür haben, die Prädiktoren im Modell zu ordnen. Typ III dagegen ist allgemeiner anwendbar. Wald-Statistiken bzw. verallgemeinerte Scorestatistiken werden anhand der Auswahl in der Gruppe "Chi-Quadrat-Statistik" berechnet.
- **Konfidenzintervalle.** Geben Sie ein Konfidenzniveau größer 50 und kleiner 100 an. Wald-Intervalle werden immer erstellt, unabhängig vom Typ der ausgewählten Chi-Quadrat-Statistik, und beruhen auf der Annahme, dass die Parameter eine asymptotische Normalverteilung aufweisen.
- **Log-Quasi-Likelihood-Funktion.** Legt das Anzeigeformat der Log-Quasi-Likelihood-Funktion fest. Die vollständige Funktion enthält einen zusätzlichen Term, der hinsichtlich der Parameterschätzungen konstant ist. Er hat keine Auswirkungen auf die Parameterschätzung und wird bei einigen Softwareprodukten nicht angezeigt.

**Drucken.** Die folgenden Ausgaben sind verfügbar:

- **Zusammenfassung der Fallverarbeitung.** Zeigt die Anzahl und den Prozentsatz der Fälle an, die in die Analyse und die Tabelle "Korrelierte Datenzusammenfassung" aufgenommen bzw. daraus ausgeschlossen werden.

- **Deskriptive Statistiken.** Zeigt eine deskriptive Statistik und Zusammenfassungsinformationen über die abhängige Variable, die Kovariaten und die Faktoren an.
- **Modellinformationen.** Zeigt den Namen des Datasets, die abhängige Variable bzw. die Ereignis- und Versuchsvariablen, die Offsetvariable, die Skalengewichtungsvariable, die Wahrscheinlichkeitsverteilung und die Verknüpfungsfunktion an.
- **Statistik für Anpassungsgüte.** Zeigt zwei Erweiterungen des Akaike-Informationskriteriums für die Modellauswahl an: Quasi-Likelihood unter Unabhängigkeitsmodellkriterium (QIC) für die Auswahl der besten Korrelationsstruktur sowie ein weiteres QIC-Maß für die Auswahl der besten Subsets an Prädiktoren.
- **Modellzusammenfassungsverstatistik.** Zeigt Tests für die Anpassungsgüte des Modells an, darunter Likelihood-Quotienten-Statistiken für den Omnibustest für die Anpassungsgüte, sowie Statistiken für Kontraste des Typs I bzw. III für jeden Effekt.
- **Parameterschätzungen.** Zeigt Parameterschätzungen und die entsprechenden Teststatistiken und Konfidenzintervalle an. Wahlweise können Sie zusätzlich zu den rohen, unbearbeiteten Parameterschätzungen auch potenzierte Parameterschätzungen anzeigen.
- **Kovarianzmatrix für Parameterschätzungen.** Zeigt die Kovarianzmatrix für die geschätzten Parameter an.
- **Korrelationsmatrix für Parameterschätzungen.** Zeigt die Korrelationsmatrix für die geschätzten Parameter an.
- **Kontrastkoeffizientenmatrizen (L-Matrizen).** Zeigt die Kontrastkoeffizienten für die Standardeffekte und für die geschätzten Randmittel an, sofern auf der Registerkarte "Geschätzte Randmittel" angefordert.
- **Allgemeine schätzbare Funktion(en).** Zeigt die Matrizen für die Generierung der Kontrastkoeffizientenmatrizen (L-Matrizen) an.
- **Iterationsverlauf.** Zeigt den Iterationsverlauf für Parameterschätzungen und Log-Likelihood an und druckt die letzte Evaluierung des Gradientenvektors und der Hesse-Matrix. Die Tabelle des Iterationsverlaufs zeigt Parameterschätzungen für jede  $n$ -te Iteration an, beginnend mit der 0-ten Iteration (Anfangsschätzungen), wobei  $n$  für den Wert des Druckintervalls steht. Wenn der Iterationsverlauf angefordert wird, wird die letzte Iteration stets angezeigt, unabhängig von  $n$ .
- **Arbeitskorrelationsmatrix.** Zeigt die Werte der Matrix an, die die Innersubjektabhängigkeiten darstellt. Ihre Struktur hängt von den Angaben auf der Registerkarte Wiederholt ab.

## Verallgemeinerte Schätzungsgleichungen: Geschätzte Randmittel

Auf dieser Registerkarte können Sie die geschätzten Randmittel für die Ebenen der Faktoren und die Faktorinteraktionen anzeigen. Des Weiteren können Sie den geschätzten Gesamtmittelwert anzeigen lassen. Geschätzte Randmittel sind für ordinal-multinomiale Modelle nicht verfügbar.

**Faktoren und Interaktionen.** Diese Liste enthält Faktoren, die auf der Registerkarte "Prädiktoren" angegeben wurden, und Interaktionen zwischen Faktoren, die auf der Registerkarte "Modell" angegeben wurden. Kovariaten sind aus dieser Liste ausgeschlossen. Terme können direkt aus dieser Liste ausgewählt werden oder mithilfe der Schaltfläche **Nach \*** zu einem Interaktionsterm kombiniert werden.

**Mittelwerte anzeigen für.** Geschätzte Mittelwerte werden für die ausgewählten Faktoren und Interaktionen zwischen Faktoren berechnet. Der Kontrast bestimmt, wie die Hypothesentests zum Vergleich der geschätzten Mittel eingerichtet werden. Für einfache Kontraste wird eine Referenzkategorie oder eine Faktorebene benötigt, mit der die anderen Elemente verglichen werden können.

- **Paarweise.** Paarweise Vergleiche werden für Kombinationen über alle Stufen der angegebenen oder implizierten Faktoren berechnet. Dies ist der einzige verfügbare Kontrast für Interaktionen zwischen Faktoren.
- **Einfach.** Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt.
- **Abweichung.** Jede Faktorstufe wird mit dem Gesamtmittelwert verglichen. Abweichungskontraste sind nicht orthogonal.

- *Differenz*. Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. Diese Kontraste werden auch als umgekehrte Helmert-Kontraste bezeichnet.
- *Helmert*. Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.
- *Wiederholt*. Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.
- *Polynomial*. Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

**Metrisch.** Geschätzte Randmittel können für die Antwort berechnet werden (auf der Grundlage der ursprünglichen Skala der abhängigen Variablen) oder für den linearen Prädiktor (auf der Grundlage der durch die Verknüpfungsfunktion transformierten abhängigen Variablen).

**Anpassung für Mehrfachvergleiche.** Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- *Bonferroni*. Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.
- *Bonferroni sequenziell*. Hierbei handelt es sich um ein sequentielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist was die Ablehnung einzelner Hypothesen angeht, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- *Sidak*. Dieses Verfahren liefert engere Grenzen als der Bonferroni-Ansatz.
- *Sidak (sequenziell)*. Hierbei handelt es sich um ein sequentielles, schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen angeht, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.

## Verallgemeinerte Schätzungsgleichungen: Speichern

Die markierten Elemente werden unter dem angegebenen Namen gespeichert. Sie können auswählen, dass vorhandene Variablen mit demselben Namen wie die neuen Variablen überschrieben werden sollen, oder Namenskonflikte durch das Anhängen von Suffixen, die die neuen Variablennamen eindeutig machen, vermeiden. .

- **Vorhergesagter Wert für Mittelwert der abhängigen Variablen.** Speichert vom Modell vorhergesagte Werte für jeden Fall in der ursprünglichen Antwortmetrik. Bei einer binomialen Antwortverteilung und einer binären abhängigen Variablen speichert die Prozedur vorhergesagte Wahrscheinlichkeiten. Bei einer multinomialen Antwortverteilung wird die Elementbeschriftung **Kumulative vorhergesagte Wahrscheinlichkeit** verwendet und die Prozedur speichert die kumulative vorhergesagte Wahrscheinlichkeit für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.
- **Untergrenze des Konfidenzintervalls für den Mittelwert der abhängigen Variablen.** Speichert die Untergrenze des Konfidenzintervalls für den Mittelwert der Antwort. Bei einer multinomialen Antwortverteilung wird die Elementbeschriftung **Untergrenze des Konfidenzintervalls für die kumulativ vorhergesagte Wahrscheinlichkeit** verwendet und die Prozedur speichert die Untergrenze für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.
- **Obergrenze des Konfidenzintervalls für den Mittelwert der Antwort.** Speichert die Obergrenze des Konfidenzintervalls für den Mittelwert der Antwort. Bei einer multinomialen Antwortverteilung wird die Elementbeschriftung **Obergrenze des Konfidenzintervalls für die kumulativ vorhergesagte Wahrscheinlichkeit** verwendet und die Prozedur speichert die Obergrenze für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

- **Vorhergesagte Kategorie.** Bei Modellen mit Binomialverteilung und binärer abhängiger Variabler bzw. mit Multinomialverteilung wird hiermit die vorhergesagte Kategorie für die einzelnen Fälle gespeichert. Diese Option ist nicht für andere Antwortverteilungen verfügbar.
- **Vorhergesagter Wert für linearen Prädiktor.** Speichert vom Modell vorhergesagte Werte für jeden Fall in der Metrik des linearen Prädiktors (transformierte Antwort über die angegebene Verknüpfungsfunktion). Bei einer multinomialen Antwortverteilung speichert die Prozedur den vorhergesagten Wert für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.
- **Geschätzter Standardfehler des vorhergesagten Werts des linearen Prädiktors.** Bei einer multinomialen Antwortverteilung speichert die Prozedur den geschätzten Standardfehler für die einzelnen Antwortkategorien, mit Ausnahme der letzten, bis zu der unter "Zu speichernde Kategorien" angegebenen Anzahl.

Die folgenden Elemente sind bei einer multinomialen Antwortverteilung nicht verfügbar.

- **Rohresiduum.** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- **Pearson-Residuum.** Die Quadratwurzel des Beitrags eines Falls zur Pearson-Chi-Quadrat-Statistik, mit dem Vorzeichen des Rohresiduums.

## Verallgemeinerte Schätzungsgleichungen: Exportieren

**Modell als Daten exportieren.** Schreibt ein Dataset im IBM SPSS Statistics-Format, das die Parameterkorrelations- oder -kovarianzmatrix enthält (mit Parameterschätzungen, Standardfehlern, Signifikanzwerten und Freiheitsgraden). Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **Teilungsvariablen.** Sofern verwendet, alle Variablen, die Aufteilungen (Splits) definieren.
- **RowType\_.** Nimmt Werte (und Wertbeschriftungen) *COV* (Kovarianzen), *CORR* (Korrelationen), *EST* (Parameterschätzungen), *SE* (Standardfehler), *SIG* (Signifikanzniveau) und *DF* (Freiheitsgrade der Stichprobenentnahme) an. Es gibt einen separaten Fall mit dem Zeilentyp *COV* (bzw. *CORR*) für jeden Modellparameter sowie einen separaten Fall für jeden der anderen Zeilentypen.
- **VarName\_.** Nimmt die Werte *P1*, *P2*, ..., die einer geordneten Liste aller geschätzten Modellparameter (mit Ausnahme der Parameter für den Maßstab oder der negativen Binomialparameter) entsprechen, für die Zeilentypen *COV* oder *CORR* an, wobei die Wertbeschriftungen den Parameterzeichenfolgen entsprechen, die in der Tabelle "Parameterschätzungen" angezeigt werden. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter (einschließlich der Maßstabswerte und der negativen binomialen Parameter) mit Variablenbeschriftungen, die den in der Tabelle "Parameterschätzungen" angezeigten Parameterzeichenfolgen entsprechen, und Werte entsprechend dem Zeilentyp.

Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzungen sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Für den Skalenparameter sind Kovarianzen, Korrelationen, Signifikanzniveau und Freiheitsgrade auf den systemdefiniert fehlenden Wert gesetzt. Wenn der Skalenparameter über Maximum Likelihood geschätzt wird, wird der Standardfehler ausgegeben, andernfalls wird der systemdefiniert fehlende Wert verwendet.

Für den negativen binomialen Parameter sind Kovarianzen, Korrelationen, Signifikanzniveau und Freiheitsgrade auf den systemdefiniert fehlenden Wert gesetzt. Wenn der negative binomiale Parameter über Maximum Likelihood geschätzt wird, wird der Standardfehler ausgegeben, andernfalls wird der systemdefiniert fehlende Wert verwendet.

Wenn Aufteilungen vorliegen, muss die Liste der Parameter über alle Splits akkumuliert sein. In einer Aufteilung können einige Parameter irrelevant sein; dies ist jedoch nicht mit redundant zu verwechseln.

Für irrelevante Parameter sind alle Kovarianzen bzw. Korrelationen, Parameterschätzungen, Standardfehler, Signifikanzniveau und Freiheitsgrade auf den systemdefiniert fehlenden Wert gesetzt.

Sie können diese Matrixdatei als Anfangswerte für die weitere Modellschätzung verwenden. Beachten Sie, dass diese Datei nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar ist, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen. Doch auch dann sollten Sie darauf achten, dass alle Parameter in dieser Matrixdatei für die Prozedur, die die Datei einliest, dieselbe Bedeutung haben.

**Modell als XML exportieren.** Speichert die Parameterschätzungen und gegebenenfalls die Parameterkovarianzmatrix (falls ausgewählt) im XML-Format (PMML). Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden. .

## Zusätzliche Funktionen beim Befehl GENLIN

Die Befehlsyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl CRITERIA können Sie Anfangswerte für Parameterschätzungen als Zahlenliste angeben.
- Mit dem Unterbefehl REPEATED können Sie eine feste Arbeitskorrelationsmatrix angeben.
- Mit dem Unterbefehl EMMEANS können Sie bei der Berechnung der geschätzten Randmittel einen anderen Wert für die Kovariaten festlegen als den Mittelwert.
- Mit dem Unterbefehl EMMEANS können Sie außerdem benutzerdefinierte polynomiale Kontraste für geschätzte Randmittel angeben.
- Mit den Schlüsselwörtern TABLES und COMPARE des Unterbefehls EMMEANS können Sie ein Subset der Faktoren, für die geschätzte Randmittel angezeigt werden, mithilfe des angegebenen Kontrasttyps vergleichen lassen.

See the *Command Syntax Reference* for complete syntax information.

## Verallgemeinerte lineare gemischte Modelle

---

Verallgemeinerte lineare gemischte Modelle erweitern das lineare Modell wie folgt:

- Das Ziel steht über eine angegebene Verknüpfungsfunktion in einer linearen Beziehung zu den Faktoren und Kovariaten.
- Das Ziel kann eine von der Normalverteilung abweichende Verteilung aufweisen.
- Es kann eine Korrelation zwischen den Beobachtungen bestehen.

Verallgemeinerte lineare gemischte Modelle decken eine breite Palette verschiedener Modelle ab, von einfacher linearer Regression bis hin zu komplexen Mehrebenenmodellen für nicht normalverteilte Longitudinaldaten.

### Beispiele

Eine Schulbehörde kann ein verallgemeinertes lineares gemischtes Modell verwenden, um zu ermitteln, ob eine experimentelle Lehrmethode die Mathematikleistungen effektiv verbessert. Studenten aus dem gleichen Klassenzimmer sollten korreliert werden, da sie von demselben Lehrer unterrichtet werden, und Klassenräume innerhalb der gleichen Schule können auch korreliert werden, so dass wir zufällige Effekte auf Schul- und Klassenstufen einbeziehen können, um verschiedene Quellen der Variabilität zu berücksichtigen.

Wissenschaftler aus der Medizinforschung können ein verallgemeinertes lineares gemischtes Modell verwenden, um zu ermitteln, ob ein neues Antikonvulsivum die Häufigkeit epileptischer Anfälle bei einem Patienten verringern kann. Messwiederholungen bei ein und demselben Patienten sind in der Regel positiv korreliert. Daher sollte ein gemischtes Modell mit einigen Zufallseffekten angemessen sein. Für das Zielfeld (Anzahl der Anfälle) werden positive ganzzahlige Werte verwendet. Daher könnte ein verallgemeinertes lineares gemischtes Modell mit einer Poisson-Verteilung und einer Log-Verknüpfung geeignet sein.

Die Geschäftsführung eines Kabelanbieters für Fernseh-, Telefon- und Internetservices kann ein verallgemeinertes lineares gemischtes Modell verwenden, um mehr über potenzielle Kunden zu erfahren. Da mögliche Antworten nominale Messniveaus aufweisen, verwendet der Unternehmensanalyst ein generalisiertes, gemischtes Modell mit einem zufälligen Intercept, um die Korrelation zwischen Antworten auf die Service-Nutzungsfragen über Servicetypen (Fernsehen, Telefon, Internet) im Rahmen einer bestimmten Umfrage Antworten der Responder zu erfassen.

Über die Registerkarte "Datenstruktur" können Sie die strukturellen Beziehungen zwischen Datensätzen in Ihrem Dataset festlegen, wenn Beobachtungen miteinander korrelieren. Wenn die Datensätze im Dataset unabhängige Beobachtungen darstellen, müssen Sie auf dieser Registerkarte nichts festlegen.

## Optionen für Effekte

### Subjekte

Die Wertekombination der angegebenen kategorialen Felder sollte die Subjekte innerhalb des Datensets eindeutig definieren. Beispiel: Ein einzelnes Feld *Patienten-ID* sollte ausreichen, um die Subjekte in einem einzelnen Krankenhaus zu definieren, doch die Kombination aus *Krankenhaus-ID* und *Patienten-ID* kann erforderlich sein, wenn die Patienten-IDs nicht krankenhausesübergreifend eindeutig sind. Bei einer Einstellung mit wiederholten Messungen werden für jedes Subjekt mehrere Beobachtungen aufgezeichnet, sodass jedes Subjekt mehrere Datensätze im Dataset belegen kann.

Ein **Subjekt** ist eine Beobachtungseinheit, die als unabhängig von anderen Subjekten betrachtet werden kann. Die Blutdruckmessungen eines Patienten in einer medizinischen Studie können beispielsweise als unabhängig von den Messungen anderer Patienten angesehen werden. Die Definition von Subjekten ist vor allem dann wichtig, wenn für jedes Subjekt Messwiederholungen durchgeführt werden und Sie die Korrelation zwischen diesen Beobachtungen analysieren möchten. So ist beispielsweise zu erwarten, dass Blutdruckmessungen bei einem bestimmten Patienten bei aufeinander folgenden Arztbesuchen miteinander korrelieren.

Alle im Dialogfeld **Variablen** als Subjekte angegebenen Felder werden dazu verwendet, Subjekte für die Kovarianzstruktur der Residuen zu definieren, und sie stellen die Liste der möglichen Felder für die Definition von Subjekten für die Kovarianzstrukturen der Zufallseffekte im Block für zufällige Effekte bereit.

### Messwiederholungen

Die hier angegebenen Felder werden verwendet, um Beobachtungswiederholungen zu kennzeichnen. So können beispielsweise mit einer einzigen Variablen für *Woche* alle 10 Wochen der Beobachtungen in einer medizinischen Studie bezeichnet werden oder die Variablen *Monat* und *Tag* können gemeinsam verwendet werden, um tägliche Beobachtungen im Verlauf eines Jahres zu bezeichnen.

## Kovarianzoptionen

### Kovarianzgruppe definieren nach

Die hier festgelegten kategorialen Felder definieren unabhängige Sets von Parametern für die Kovarianzen der wiederholten Effekte; einen für jede Kategorie, die durch die Kreuzklassifikation der Gruppierungsfelder definiert werden. Alle Subjekte weisen denselben Kovariantyp auf. Subjekte innerhalb derselben Kovarianzgruppierung weisen dieselben Werte für die Parameter auf.

### Kovariantyp bei Messwiederholung

Hiermit wird die Kovarianzstruktur für die Residuen angegeben. Unterschiedliche Kovarianzoptionen sind basierend auf dem ausgewählten **Kovariantyp bei Messwiederholung** verfügbar. Die folgenden Strukturen sind verfügbar:

- Autoregressiv der ersten Ordnung (AR1)
- Direktes Produkt AR1(UN\_AR1)
- Direktes Produkt, unstrukturiert (UN\_UN)
- Direktprodukt, zusammengesetzte Symmetrie (UN\_CS)
- Heterogene zusammengesetzte Symmetrie (CSH)
- Heterogen autoregressiv (ARH1)

- Autoregressiv mit gleitendem Durchschnitt (1,1) (ARMA11)
- Zusammengesetzt symmetrisch (ZS)
- Diagonal
- Skalierte Identität
- Toeplitz
- Unstrukturiert
- Varianzkomponenten
- Räumlich: Exponent
- Räumlich: exponentiell
- Räumlich: gaußsch
- Räumlich: linear
- Räumlich: linear logistisch
- Räumlich: kugelförmig

### Kronecker-Messungen

Wählen Sie Variablen aus, die die Subjektstruktur für Kronecker-Kovarianz-Messungen angeben und bestimmen, wie die Messfehler korreliert werden. Das Feld ist nur verfügbar, wenn eine der folgenden Optionen unter **Kovarianztyp bei wiederholter Kovarianz** ausgewählt ist:

- Direktes Produkt AR1(UN\_AR1)
- Direktes Produkt, unstrukturiert (UN\_UN)
- Direktprodukt, zusammengesetzte Symmetrie (UN\_CS)

### Koordinaten für räumliche Kovarianz

Die Variablen in dieser Liste geben die Koordinaten der Beobachtungswiederholungen an, wenn einer der räumlichen Kovarianztypen als Kovarianztyp bei Messwiederholung ausgewählt wird.

Weitere Informationen finden Sie in „Kovarianzstrukturen“ auf Seite 107.

### Pseudo R<sup>2</sup> -Maße

Pseudo-R<sup>2</sup> -Maße und der Korrelationskoeffizient innerhalb der Klasse werden in die GLMM-Ausgabe eingeschlossen (falls zutreffend). Die Pseudo-R<sup>2</sup> -Maße basieren vollständig auf den endgültigen Schätzungen und werden nach Abschluss der Schätzung erstellt. Der Bestimmungskoeffizient R<sup>2</sup> ist eine häufig gemeldete Statistik, da er den Anteil der Varianz darstellt, der durch ein lineares Modell erklärt wird. Der Intra-Klasse-Korrelationskoeffizient (ICC) ist eine verwandte Statistik, die den Anteil der Varianz quantifiziert, der durch einen Gruppierungsfaktor (Zufallsfaktor) in mehrstufigen/hierarchischen Daten erklärt wird.

## Erzeugen eines verallgemeinerten linearen gemischten Modells

Für diese Funktion SPSS Statistics ist die Standard Edition oder die Option "Advanced Statistics" erforderlich.

Wählen Sie Folgendes in den Menüs aus:

### Analysieren > Gemischte Modelle > Verallgemeinerte lineare Modelle...

1. Legen Sie die Subjektstruktur Ihres Datasets auf der Registerkarte **Datenstruktur** fest. Die Registerkarte stellt Optionen zur Angabe der Modelleffekte und des wiederholten Kovarianztyps bereit.
2. Auf der Registerkarte **Felder und Effekte** muss ein einzelnes Ziel vorhanden sein, das jedes beliebige Messniveau oder eine Festlegung für Ereignisse/Versuche aufweisen kann, in welchem Fall die Festlegungen für die Ereignisse und Versuche kontinuierlich sein müssen. Optional können Sie seine Verteilung und Verknüpfungsfunktion, die festen Effekte und etwaige Blöcke mit zufälligen Effekten, Offset oder Analysegewichtungen angeben.
3. Klicken Sie auf **Erstellungsoptionen**, um optionale Erstellungseinstellungen anzugeben.

4. Klicken Sie auf **Modelloptionen**, um Scores im aktiven Dataset zu speichern und das Modell an eine externe Datei zu exportieren.
5. Klicken Sie auf **Ausführen**, um die Prozedur auszuführen und die Modellobjekte zu erstellen.

## Ziel

Mit diesen Einstellungen werden das Ziel, seine Verteilung und seine Beziehung zu den Prädiktoren über die Verknüpfungsfunktion definiert.

**Ziel.** Das Ziel muss angegeben werden. Es kann jedes beliebige Messniveau aufweisen. Durch das Messniveau des Ziels wird die Menge der jeweils geeigneten Verteilungen und Verknüpfungsfunktionen eingegrenzt.

- **Anzahl der Versuche als Nenner verwenden.** Wenn die Zielantwort eine Anzahl von Ereignissen ist, die in einer Menge von Versuchen eintreten, enthält das Zielfeld die Anzahl der Ereignisse und Sie können ein zusätzliches Feld auswählen, das die Anzahl der Versuche enthält. Beim Testen eines neuen Pestizids können Sie beispielsweise Stichproben von Ameisen verschiedenen Konzentrationen des Schädlingsbekämpfungsmittels aussetzen. Zeichnen Sie dabei die Anzahl der vernichteten Ameisen und die Anzahl der dem Mittel ausgesetzten Ameisen auf. In diesem Fall sollte das Feld, in dem die Zahl der vernichteten Ameisen aufgezeichnet wird, als Zielfeld (Ereignisse) und das Feld, in dem die Zahl der Ameisen in den einzelnen Stichproben aufgezeichnet wird, als Feld für die Versuche festgelegt werden. Wenn die Zahl der Ameisen in den einzelnen Stichproben gleich ist, kann die Anzahl der Versuche mit einem festen Wert angegeben werden.

Die Anzahl der Versuche sollte größer oder gleich der Anzahl der Ereignisse für die einzelnen Datensätze sein. Bei den Ereignissen sollte es sich um nicht negative Ganzzahlen und bei den Versuchen um positive Ganzzahlen handeln.

- **Referenzkategorie anpassen.** Bei einem kategorialen Ziel können Sie die Referenzkategorie auswählen. Dies kann sich auf bestimmte Ausgaben, wie beispielsweise Parameterschätzungen, auswirken, sollte jedoch nicht die Anpassungsgüte des Modells verändern. Beispiel: Angenommen, Ihr Ziel kann die Werte 0, 1 und 2 annehmen. In diesem Fall verwendet die Prozedur standardmäßig die letzte Kategorie (die mit dem höchsten Wert), also 2, als Referenzkategorie. In dieser Situation sollten Parameterschätzungen so interpretiert werden, dass sie die Wahrscheinlichkeit der Kategorie 0 oder 1 *relativ* auf die Wahrscheinlichkeit der Kategorie 2 beziehen. Wenn Sie eine angepasste Kategorie angeben und Ihr Ziel definierte Beschriftungen hat, können Sie die Referenzkategorie festlegen, indem Sie einen Wert aus der Liste auswählen. Dies kann nützlich sein, wenn Sie bei der Festlegung eines Modells nicht mehr wissen, wie genau ein bestimmtes Feld codiert war.

**Zielverteilung und Beziehung (Verknüpfung) mit dem linearen Modell.** Angesichts der Werte der Prädiktoren geht das Modell davon aus, dass die Verteilung der Werte des Ziels der angegebenen Form folgt und dass die Zielwerte über die angegebene Verknüpfungsfunktion in einer linearen Beziehung zu den Prädiktoren stehen. Verknüpfungen für mehrere häufig verwendete Modelle werden bereitgestellt; Sie können auch eine **benutzerdefinierte** Einstellung auswählen, wenn es eine bestimmte Kombination einer Verteilung und einer Verknüpfungsfunktion gibt, die Sie anpassen möchten und die nicht in der Liste der Schnellverfahren enthalten ist.

- **Lineares Modell.** Gibt eine Normalverteilung mit einer Identitätsverknüpfung an, was nützlich ist, wenn sich das Ziel mit einer linearen Regression oder einem ANOVA-Modell vorhersagen lässt.
- **Gamma-Regression.** Gibt eine Gammaverteilung mit einer Log-Verknüpfung an, die eingesetzt werden sollte, wenn das Ziel ausschließlich positive Werte enthält und eine Verzerrung hin zu größeren Werten aufweist.
- **Loglinear.** Gibt eine Poisson-Verteilung mit einer Log-Verknüpfung an, die eingesetzt werden sollte, wenn das Ziel eine Anzahl an Vorkommen in einem festen Zeitraum darstellt.
- **Negative binomiale Regression.** Gibt eine negative Binomialverteilung mit einer Log-Verknüpfung an, die eingesetzt werden sollte, wenn Ziel und Nenner die Anzahl der Versuche darstellen, die erforderlich sind um  $k$  Erfolge zu beobachten.
- **Multinomiale logistische Regression.** Gibt eine Multinomialverteilung an, die eingesetzt werden sollte, wenn es sich bei dem Ziel um eine Antwort mit mehreren Kategorien handelt. Verwendet entweder eine

Verknüpfung vom Typ "Logit (kumulativ)" (ordinale Ergebnisse) oder eine Verknüpfung vom Typ "Logit (verallgemeinert)" (nominale Antworten mit mehreren Kategorien).

- **Binäre logistische Regression.** Gibt eine Binomialverteilung mit einer Logit-Verknüpfung an, die eingesetzt werden sollte, wenn es sich bei dem Ziel um eine Binärantwort handelt, die durch ein logistisches Regressionsmodell vorhergesagt wird.
- **Binär Probit.** Gibt eine Binomialverteilung mit einer Probit-Verknüpfung an, die eingesetzt werden sollte, wenn es sich bei dem Ziel um eine Binärantwort handelt, der eine Normalverteilung zugrunde liegt.
- **Intervallzensiertes Überleben.** Gibt eine Binomialverteilung mit einer Verknüpfung vom Typ "Log-Log komplementär" an, die sinnvoll für Überlebensanalysen ist, bei denen einige Beobachtungen kein Beendigungsereignis aufweisen.

#### Verteilung

Diese Auswahl gibt die Verteilung des Ziels an. Die Möglichkeit einer anderen Verteilung als "Normal" und einer anderen Verknüpfungsfunktion als "Identität" ist die wichtigste Verbesserung des verallgemeinerten linearen gemischten Modells gegenüber dem linearen gemischten Modell. Es gibt zahlreiche mögliche Kombinationen aus Verteilung und Verknüpfungsfunktion und es können mehrere davon für das jeweils vorliegende Dataset geeignet sein. Daher können Sie sich in Ihrer Wahl durch theoretische Vorüberlegungen leiten lassen oder davon, welche Kombination am besten zu passen scheint.

- **Binomial.** Diese Verteilung ist nur für Ziele geeignet, die eine binäre Antwort oder eine Anzahl von Ereignissen repräsentieren.
- **Gamma.** Diese Verteilung eignet sich für Ziele mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Wenn ein Datenwert kleiner-gleich 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Invers normal.** Diese Verteilung eignet sich für Ziele mit positiven Skalenwerten, die in Richtung größerer positiver Werte verzerrt sind. Wenn ein Datenwert kleiner-gleich 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.
- **Multinomial.** Diese Verteilung eignet sich für Ziele, die eine Antwort mit mehreren Kategorien repräsentieren. Die Form des Modells hängt vom Messniveau des Ziels ab.

Ein **nominales** Ziel führt zu einem nominal multinomialen Modell, bei dem für jede Zielkategorie mit Ausnahme der Referenzkategorie ein separates Set an Modellparametern geschätzt wird. Die Parameterschätzungen für einen Prädiktor zeigen jeweils die Beziehung zwischen dem betreffenden Prädiktor und der Wahrscheinlichkeit der einzelnen Kategorien des Ziels, relativ zur Referenzkategorie.

Ein **ordinales** Ziel führt zu einem ordinal multinomialen Modell, bei dem der herkömmliche konstante Term mit einem Set **Schwellenparameter** ersetzt wird, die in Relation zur kumulativen Wahrscheinlichkeit der Zielkategorien stehen.

- **Negativ binomial.** Bei der negativen binomialen Regression wird eine negative Binomialverteilung mit einer Log-Verknüpfung verwendet, die eingesetzt werden sollte, wenn es sich bei dem Ziel um eine Zählung von Vorkommnissen mit hoher Varianz handelt.
- **Normal.** Diese Option eignet sich für stetige Ziele, deren Werte eine symmetrische, glockenförmige Verteilung um einen Mittelwert aufweisen.
- **Poisson.** Diese Verteilung lässt sich als Anzahl der Vorkommen eines untersuchten Ereignisses in einem festen Zeitraum betrachten und eignet sich für Variablen mit nicht negativen ganzzahligen Werten. Wenn ein Datenwert keine Ganzzahl oder kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.

#### Verknüpfungsfunktionen

Die Verknüpfungsfunktion ist eine Transformation des Ziels, die eine Schätzung des Modells ermöglicht. Die folgenden Funktionen sind verfügbar:

- **Identität.**  $f(x)=x$ . Das Ziel ist nicht transformiert. Diese Verknüpfung kann abgesehen von der Multinomialverteilung mit jeder beliebigen Verteilung verwendet werden.

- **Komplementäres Protokoll-Protokoll.**  $f(x)=\log(-\log(1-x))$ . Nur für die Binomialverteilung oder Multinomialverteilung geeignet.
- **Cauchit.**  $f(x) = \tan(\pi (x - 0.5))$ . Nur für die Binomialverteilung oder Multinomialverteilung geeignet.
- **Log.**  $f(x) = \log(x)$ . Diese Verknüpfung kann abgesehen von der Multinomialverteilung mit jeder beliebigen Verteilung verwendet werden.
- **Protokollergänz.**  $f(x)=\log(1-x)$ . Nur für die Binomialverteilung geeignet.
- **Logit.**  $f(x) = \log(x/(1-x))$ . Nur für die Binomialverteilung oder Multinomialverteilung geeignet.
- **Negatives Protokoll-Protokoll.**  $f(x) = -\log (-\log (x))$ . Nur für die Binomialverteilung oder Multinomialverteilung geeignet.
- **Probit.**  $f(x) = \Phi^{-1}(x)$ , wobei  $\Phi^{-1}$  die umgekehrte kumulative Standardnormalverteilungsfunktion ist. Nur für die Binomialverteilung oder Multinomialverteilung geeignet.
- **Powert.**  $f(x) = x^\alpha$ , wenn  $\alpha \neq 0$  ist.  $f(x) = \log (x)$ , wenn  $\alpha = 0$  ist.  $\alpha$  ist die erforderliche Zahlenangabe und muss eine reelle Zahl sein. Diese Verknüpfung kann abgesehen von der Multinomialverteilung mit jeder beliebigen Verteilung verwendet werden.

## Feste Effekte

Faktoren mit festen Effekten werden im Allgemeinen als Felder betrachtet, deren relevante Werte alle im Dataset dargestellt werden und für das Scoring verwendet werden können. Standardmäßig werden Felder mit der vordefinierten Eingaberolle, die nicht an anderer Stelle des Dialogs angegeben sind, in den Bereich des Modells eingegeben, der feste Effekte aufweist. Kategoriale Felder (nominal und ordinal) werden als Faktoren im Modell verwendet und stetige Felder werden als Kovariaten verwendet.

Geben Sie Effekte in das Modell ein, indem Sie ein oder mehrere Felder in der Quellenliste auswählen und sie in die Liste der Effekte ziehen. Welche Art von Effekt erstellt wird, hängt davon ab, auf welchem Hotspot Sie die Auswahl ablegen.

- **Haupt.** Die abgelegten Felder werden unten in der Liste der Effekte als separate Haupteffekte angezeigt.
- **Zweiweg.** Alle möglichen Paare der abgelegten Felder werden unten in der Liste der Effekte als Zweiwegeinteraktionen angezeigt.
- **Dreiweg.** Alle möglichen Dreiergruppen der abgelegten Felder werden unten in der Liste der Effekte als Dreiwegeinteraktionen angezeigt.
- **\***. Die Kombination aller gelöschten Felder wird als eine einzelne Interaktion am unteren Rand der Effektliste angezeigt.

Schaltflächen rechts neben dem Effektgenerator ermöglichen die Ausführung verschiedener Aktionen:

*Tabelle 1. Beschreibung der Schaltflächen des Effektgenerators*

Symbol	Beschreibung
	Löschen von Termen aus dem Modell mit festen Effekten durch Auswahl der Terme, die Sie löschen möchten, und durch Klicken auf die Schaltfläche zum Löschen.
	Umordnen der Terme innerhalb des Modells mit festen Effekten, durch Auswahl der Terme, die Sie umordnen möchten, und durch Klicken auf den Aufwärts- oder Abwärtspeil.
	Hinzufügen von verschachtelten Termen zum Modell mithilfe des Dialogfelds zum „Hinzufügen eines benutzerdefinierten Terms“ auf Seite 59 durch Klicken auf die Schaltfläche "Benutzerdefinierten Term hinzufügen".

**Konstanten Term einschließen.** Der konstante Term wird normalerweise in das Modell eingeschlossen. Wenn anzunehmen ist, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

## Hinzufügen eines benutzerdefinierten Terms

In dieser Prozedur können Sie verschachtelte Terme für Ihr Modell erstellen. Verschachtelte Terme sind nützlich, um den Effekt eines Faktors oder einer Kovariaten zu modellieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. So kann beispielsweise eine Lebensmittelkette den Ausgabengewohnheiten ihrer Kunden an mehreren Filialstandorten folgen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

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

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

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

Erstellen eines verschachtelten Terms

1. Wählen Sie einen Faktor oder eine Kovariaten aus, der bzw. die in einem anderen Faktor verschachtelt ist, und klicken Sie auf die Pfeilschaltfläche.
2. Klicken Sie auf **(Innerhalb)**.
3. Wählen Sie den Faktor aus, in dem der vorherige Faktor bzw. die vorherige Kovariaten verschachtelt ist, und klicken Sie dann auf die Pfeilschaltfläche.
4. Klicken Sie auf **Term hinzufügen**.

Optional können Sie Interaktionseffekte einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

## Zufällige Effekte

Zufallseffektfaktoren sind Felder, deren Wert in der Datendatei als zufällige Stichprobe aus einer größeren Gesamtheit von Werten betrachtet werden kann. Sie helfen bei der Erklärung von übermäßiger Variabilität beim Ziel. Wenn Sie auf der Registerkarte "Datenstruktur" mehrere Objekte ausgewählt haben, wird standardmäßig für jedes Subjekt, das über das innerste Subjekt hinausgeht, ein Block mit zufälligen Effekten erstellt. Wenn Sie beispielsweise auf der Registerkarte "Datenstruktur" als Subjekte "Schule", "Klasse" und "Schüler" ausgewählt haben, werden automatisch die folgenden Blöcke für zufällige Effekte erstellt:

- Zufälliger Effekt 1: Subjekt ist Schule (ohne Effekte, nur konstanter Term)
- Zufälliger Effekt 2: Subjekt ist Schule \* Klasse (keine Effekte, nur konstanter Term)

Sie können wie folgt mit Blöcken mit zufälligen Effekten arbeiten:

1. Klicken Sie auf **Block hinzufügen...**, um einen neuen Block hinzuzufügen. Daraufhin wird das Dialogfeld „Block für zufällige Effekte“ auf Seite 59 geöffnet.
2. Um einen vorhandenen Block zu bearbeiten, wählen Sie den Block aus, den Sie bearbeiten möchten, und klicken Sie auf **Block bearbeiten...**. Daraufhin wird das Dialogfeld „Block für zufällige Effekte“ auf Seite 59 geöffnet.
3. Um einen oder mehrere Blöcke zu löschen, wählen Sie die betreffenden Blöcke aus und klicken Sie auf die Löschschriftfläche.

## Block für zufällige Effekte

Geben Sie Effekte in das Modell ein, indem Sie mindestens ein Feld in der Quellenliste auswählen und es in der Liste **Effekterstellung** hinzufügen.

Welche Art von Effekt erstellt wird, hängt davon ab, welchen **Typ** sie auswählen. Kategoriale Felder (nominal und ordinal) werden als Faktoren im Modell verwendet und stetige Felder werden als Kovariaten verwendet.

### **Interaktion**

Die Kombination aller Felder wird unten in der Liste der Effekte als Einzelinteraktion angezeigt.

### **Haupteffekte**

Die abgelegten Felder werden unten in der Liste der Effekte als separate Haupteffekte angezeigt.

### **Alle 2-Wege**

Alle möglichen Paare der abgelegten Felder werden unten in der Liste der Effekte als Zweigeinteraktionen angezeigt.

### **Alle 3-Wege**

Alle möglichen Dreiergruppen der abgelegten Felder werden unten in der Liste der Effekte als Dreigeinteraktionen angezeigt.

### **Alle 4-Wege**

Alle möglichen Dreiergruppen der abgelegten Felder werden unten in der Liste der Effekte als Viergeinteraktionen angezeigt.

### **Alle 5-Wege**

Alle möglichen Dreiergruppen der abgelegten Felder werden unten in der Liste der Effekte als Fünfgeinteraktionen angezeigt.

### **Konstanten Term einschließen**

Der konstante Term ist standardmäßig nicht im Modell mit zufälligen Effekten enthalten. Wenn anzunehmen ist, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

### **Parametervorhersagen für diesen Block anzeigen**

Gibt an, dass die Parameterschätzungen für zufällige Effekte angezeigt werden sollen.

### **Subjektkombination**

Hiermit können Sie eine Auswahl aus vorab festgelegten Kombinationen von Subjekten mit zufälligen Effekten im Dialogfeld "Variablen" treffen. Wenn beispielsweise im Dialogfeld "Variablen" die Subjekte *Schule*, *Klasse* und *Schüler* (in dieser Reihenfolge) definiert sind, enthält die Dropdown-Liste für die Subjektkombinationen folgende Optionen: **Keines**, **Schule**, **Schule \* Klasse** und **Schule \* Klasse \* Schüler**.

### **Kovarianztyp für zufällige Effekte**

Hiermit wird die Kovarianzstruktur für die Residuen angegeben. Die folgenden Strukturen sind verfügbar:

- Autoregressiv der ersten Ordnung (AR1)
- Autoregressiv mit gleitendem Durchschnitt (1,1) (ARMA11)
- Zusammengesetzt symmetrisch
- Diagonal
- Skalierte Identität
- Toeplitz
- Unstrukturiert
- Varianzkomponenten

### **Kovarianzgruppe definieren nach**

Die hier festgelegten kategorialen Felder definieren unabhängige Sets von Parametern für die Kovarianzen der zufälligen Effekte, nämlich eines für jede durch die Kreuzklassifikation der Gruppierungsfelder definierte Kategorie. Für die einzelnen Blöcke für zufällige Effekte können unterschiedliche Mengen von Gruppierungsfeldern festgelegt werden. Alle Subjekte weisen denselben Kovarianztyp auf. Subjekte innerhalb derselben Kovarianzgruppierung weisen dieselben Werte für die Parameter auf.

## Gewichtung und Offset

**Analysegewichtung.** Der Skalenparameter ist ein geschätzter Modellparameter, der mit der Varianz der Antwort zusammenhängt. Die Analysegewichtungen sind "bekannte" Werte, die sich zwischen den einzelnen Beobachtungen unterscheiden können. Wenn das Feld 'Analysegewichtung' angegeben ist, wird der Skalenparameter, der mit der Varianz der Antwort zusammenhängt, für jede Beobachtung durch die Werte für die Analysegewichtung geteilt. Datensätze, deren Analysegewichtungswerte kleiner-gleich 0 sind oder fehlen, werden in der Analyse nicht verwendet.

**Offset.** Der Term "Offset" ist ein "struktureller" Prädiktor. Sein Koeffizient wird nicht vom Modell geschätzt, sondern es wird davon ausgegangen, dass er den Wert 1 aufweist. Daher werden die Werte des Offsets einfach zum linearen Prädiktor des Ziels addiert. Dies ist besonders nützlich bei Poisson-Regressionsmodellen, bei denen die verschiedenen Fälle dem relevanten Ereignis unterschiedlich stark ausgesetzt sein können.

Beispielsweise gibt es bei der Modellierung der Unfallraten für einzelne Fahrer einen wichtigen Unterschied zwischen einem Fahrer, der in 3 Jahren Fahrpraxis einen Unfall verursacht hat und einem Fahrer, der in 25 Jahren einen Unfall verursacht hat. Die Anzahl der Unfälle kann als Poisson- oder negative Binomialantwort mit einer Protokollverknüpfung modelliert werden, wenn der natürliche Logarithmus der Fahrpraxis des Fahrers als Offset-Term eingeschlossen wird.

Andere Kombinationen der Verteilung und Verknüpfungstypen machen andere Transformationen der Offsetvariable erforderlich.

## Allgemeine Erstellungsoptionen

Mit dieser Auswahl werden einige erweiterte Kriterien angegeben, die für die Erstellung des Modells verwendet werden.

### Sortierreihenfolge

Mit diesen Steuerelementen wird die Reihenfolge der Kategorien für die Ziele und Faktoren (kategoriale Eingaben) festgelegt, um die "letzte" Kategorie zu ermitteln. Die Einstellung für die Sortierreihenfolge des Ziels wird ignoriert, wenn das Ziel nicht kategorial ist oder wenn in den Einstellungen für „Ziel“ [auf Seite 56](#) eine benutzerdefinierte Referenzkategorie angegeben wurde.

### Stoppregeln

Sie können die maximale Anzahl an Iterationen angeben, die im Algorithmus vorgenommen werden. Der Algorithmus verwendet einen doppelt iterativen Prozess, der aus einer inneren und einer äußeren Schleife besteht. Der Wert, der als maximale Anzahl Iterationen angegeben ist, gilt für beide Schleifen. Geben Sie eine nicht negative Ganzzahl an. Der Standardwert beträgt 100.

### Einstellungen nach der Schätzung

Mit diesen Einstellungen wird festgelegt, wie ein Teil der Modellausgabe für die Anzeige berechnet wird.

### Konfidenzniveau (%)

Das Konfidenzniveau wird zur Berechnung der Intervallschätzungen der Modellkoeffizienten verwendet. Geben Sie einen Wert größer als 0 und kleiner als 100 ein. Der Standardwert ist 3.

### Freiheitsgrade

Damit wird angegeben, wie Freiheitsgrade für Signifikanztests berechnet werden. Wählen Sie die Option **Residualmethode** aus, wenn Ihre Stichprobe groß genug ist oder die Daten ausgewogen sind oder das Modell einen einfacheren Kovarianztyp verwendet (zum Beispiel "Skalierte Identität" oder "Diagonal"). Hierbei handelt es sich um die Standardeinstellung. Wählen Sie die Option **Satterthwaite-Approximation** aus, wenn Ihre Stichprobe klein ist oder die Daten unausgewogen sind oder das Modell einen komplizierten Kovarianztyp verwendet (zum Beispiel "Unstrukturiert"). Wählen Sie die Option **Kenward-Roger-Approximation** aus, wenn Ihre Stichprobengröße klein ist und Sie ein eingeschränktes REML-Modell (Restricted Maximum Likelihood) haben.

### Tests für feste Effekte und Koeffizienten

Dies ist die Methode zur Berechnung der Kovarianzmatrix der Parameterschätzungen. Wählen Sie die robuste Schätzung, wenn Sie befürchten, dass die Modellannahmen verletzt sein könnten.

## Schätzung

Der Modellerstellungsalgorithmus verwendet einen doppelt iterativen Prozess, der aus einer inneren und einer äußeren Schleife besteht. Die folgenden Einstellungen gelten für die innere Schleife.

### Parameterkonvergenz.

Es wird angenommen, dass eine Konvergenz vorliegt, wenn die maximale absolute oder maximale relative Änderung in den Parameterschätzungen kleiner als der angegebene Wert ist, der nicht negativ sein darf. Wenn der angegebene Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

### Log-Likelihood-Konvergenz.

Es wird angenommen, dass eine Konvergenz vorliegt, wenn die absolute oder relative Änderung in der Log-Likelihood-Funktion kleiner als der angegebene Wert ist, der nicht negativ sein darf. Wenn der angegebene Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

### Hesse-Konvergenz.

Für die Spezifikation **Absolut** wird angenommen, dass eine Konvergenz vorliegt, wenn eine Statistik auf der Basis der Hesse-Matrix kleiner als der angegebene Wert ist. Für die Spezifikation **Relativ** wird angenommen, dass eine Konvergenz vorliegt, wenn die Statistik kleiner als das Produkt aus dem angegebenen Wert und dem absoluten Wert der Log-Likelihood ist. Wenn der angegebene Wert gleich 0 ist, wird dieses Kriterium nicht angewendet.

### Maximale Fisher-Scoring-Schritte.

Geben Sie eine nicht negative Ganzzahl an. Der Wert 0 gibt die Newton-Raphson-Methode an. Werte größer als 0 geben die Verwendung des Fisher-Scoring-Algorithmus bis Iterationsnummer  $n$  an, wobei  $n$  die angegebene Ganzzahl ist. Anschließend wird Newton-Raphson verwendet.

### Toleranz für Prüfung auf Singularität.

Dieser Wert wird als Toleranz bei der Prüfung auf Singularität verwendet. Geben Sie einen positiven Wert ein.

**Anmerkung:** Standardmäßig wird "Parameterkonvergenz" verwendet. Dabei wird die maximale **absolute** Änderung mit einer Toleranz von  $1E-6$  geprüft. Diese Einstellung kann zu Ergebnissen führen, die sich von den Ergebnissen unterscheiden, die in Versionen vor Version 22 erhalten wurden. Um Ergebnisse der Versionen vor Version 22 zu reproduzieren, verwenden Sie **Relativ** als Wert für das Kriterium "Parameterkonvergenz" und übernehmen Sie den Standardtoleranzwert von  $1E-6$ .

## Geschätzte Mittelwerte

Auf dieser Registerkarte können Sie die geschätzten Randmittel für die Ebenen der Faktoren und die Faktorinteraktionen anzeigen. Geschätzte Randmittel sind für multinomiale Modelle nicht verfügbar.

**Terme.** Die Modellterme in den festen Effekten, die gänzlich aus kategorialen Feldern bestehen, werden hier aufgeführt. Markieren Sie jeden Term, für den das Modell geschätzte Randmittel erstellen soll.

- **Kontrasttyp.** Damit wird der Kontrasttyp angegeben, der für die Stufen des Kontrastfelds verwendet wird. Wenn **Keine** ausgewählt ist, werden keine Kontraste erstellt. **Paarweise** führt zu paarweisen Vergleichen für Kombinationen über alle Stufen der angegebenen Faktoren. Dies ist der einzige verfügbare Kontrast für Interaktionen zwischen Faktoren. Bei Kontrasten vom Typ **Abweichung** wird jede Faktorstufe mit dem Gesamtmittelwert verglichen. **Einfache** Kontraste vergleichen jede Stufe des Faktors, mit Ausnahme der letzten, mit der letzten Stufe. Die "letzte" Stufe wird durch die Sortierreihenfolge für Faktoren bestimmt, die in den Erstellungsoptionen festgelegt ist. Beachten Sie, dass keiner dieser Kontrasttypen orthogonal ist.
- **Kontrastfeld.** Hier wird ein Faktor angegeben, dessen Stufen mithilfe des ausgewählten Kontrasttyps verglichen werden. Wenn **Keiner** als Kontrasttyp ausgewählt ist, kann (oder muss) kein Kontrastfeld ausgewählt werden.

**Stetige Felder.** Die aufgeführten stetigen Felder werden aus den Termen in den festen Effekten extrahiert, bei denen stetige Felder verwendet werden. Bei der Berechnung der geschätzten Randmittel sind die Kovariaten auf die angegebenen Werte festgelegt. Wählen Sie den Mittelwert aus oder geben Sie einen benutzerdefinierten Wert an.

**Geschätzte Mittelwerte anzeigen als.** Damit wird angegeben, ob geschätzte Randmittel anhand der ursprünglichen Skala des Ziels oder anhand der Transformation einer Verknüpfungsfunktion berechnet werden sollen. **Ursprüngliche Zielskala** berechnet geschätzte Randmittel für das Ziel. Beachten Sie: Wenn das Ziel mithilfe der Option für Ereignisse/Versuche angegeben wird, werden hiermit die geschätzten Randmittel für das Verhältnis Ereignisse/Versuche ausgegeben und nicht für die Anzahl der Ereignisse. **Transformation einer Verknüpfungsfunktion** berechnet geschätzte Randmittel für den linearen Prädiktor.

**Anpassen für Mehrfachvergleiche mithilfe von.** Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. Damit können Sie die Anpassungsmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- *Bonferroni sequenziell.* Hierbei handelt es sich um ein sequentielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist was die Ablehnung einzelner Hypothesen angeht, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- *Sidak (sequenziell).* Hierbei handelt es sich um ein sequentielles, schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen angeht, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.

Die Methode der geringsten signifikanten Differenz ist weniger konservativ als die sequenzielle Sidak-Methode, die wiederum weniger konservativ ist als die sequenzielle Bonferroni-Methode. Bei der Methode der geringsten signifikanten Differenz werden also mindestens so viele einzelne Hypothesen verworfen wie bei der sequenziellen Sidak-Methode, bei der wiederum mindestens so viele einzelne Hypothesen verworfen werden wie bei der sequenziellen Bonferroni-Methode.

## Speichern

Gekennzeichnete Elemente werden mit dem angegebenen Namen gespeichert; Konflikte mit bestehenden Feldnamen sind nicht zulässig.

**Vorhergesagte Werte-** Speichert den vorhergesagten Wert des Ziels. Der Standardfeldname lautet *Vorhergesagter Wert*.

**Vorhergesagte Wahrscheinlichkeit für kategoriale Ziele-** Wenn das Ziel kategorial ist, speichert dieses Schlüsselwort die vorhergesagten Wahrscheinlichkeiten der ersten  $n$  Kategorien bis zu dem als **Maximal zu speichernde Kategorien** angegebenen Wert. Die berechneten Werte sind kumulative Wahrscheinlichkeiten für ordinale Ziele. Der Standardstammname lautet *PredictedProbability*. Um die vorhergesagte Wahrscheinlichkeit der vorhergesagten Kategorie zu speichern, speichern Sie die Konfidenz (siehe unten).

**Konfidenzintervalle-** Speichert Ober- und Untergrenzen des Konfidenzintervalls für den vorhergesagten Wert oder die vorhergesagte Wahrscheinlichkeit. Bei allen Verteilungen mit Ausnahme der Multinomialverteilung werden damit zwei Variablen erstellt. Der Standardname für den Stamm lautet *CI*, mit *\_Lower* und *\_Upper* als Suffixe.

Bei der multinomialen Verteilung und einem nominalen Ziel wird für jede abhängige Variablenkategorie ein Feld erstellt. Damit wird die Unter- und Obergrenze der vorhergesagten Wahrscheinlichkeiten für die ersten  $n$  Kategorien bis zu dem unter **Maximal zu speichernde Kategorien** angegebenen Wert gespeichert. Der Standardname für den Stamm lautet *CI* und die Standardfeldnamen lauten *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2* usw. (entsprechend der Reihenfolge der Zielkategorien).

Bei der multinomialen Verteilung und einem ordinalen Ziel wird für jede abhängige Variablenkategorie mit Ausnahme der letzten ein Feld erstellt (weitere Informationen finden Sie in „Allgemeine Erstellungsoptionen“ auf Seite 61). Damit wird die Unter- und die Obergrenze der kumulativen vorhergesagten

Wahrscheinlichkeit für die ersten  $n$  Kategorien (maximal bis einschließlich der vorletzten Kategorie) und bis zu dem unter **Maximal zu speichernde Kategorien** angegebenen Wert gespeichert. Der Standardname für den Stamm lautet *CI* und die Standardfeldnamen lauten *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2* usw. (entsprechend der Reihenfolge der Zielkategorien).

**Pearson-Residuen-** Speichert das Pearson-Residuum für jeden Datensatz, das in den Diagnosen nach der Schätzung für die Anpassungsgüte des Modells verwendet wird. Der Standardfeldname lautet *PearsonResidual*.

**Konfidenzen-** Speichert die Konfidenz im vorhergesagten Wert für das kategoriale Ziel. Die berechnete Konfidenz kann auf der Wahrscheinlichkeit des vorhergesagten Werts (der höchsten vorhergesagten Wahrscheinlichkeit) oder auf der Differenz zwischen der höchsten vorhergesagten Wahrscheinlichkeit und der zweithöchsten vorhergesagten Wahrscheinlichkeit beruhen. Der Standardfeldname lautet *Konfidenz*.

## Exportieren

**Modell exportieren.** Schreibt das Modell in eine externe *.zip*-Datei. Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden. Weitere Informationen finden Sie im Thema [Scoring-Assistent](#). Geben Sie einen eindeutigen, gültigen Dateinamen an. Wenn die Dateispezifikation eine bestehende Datei angibt, wird diese Datei überschrieben.

### Ausgabe exportieren

**EBLUPS exportieren.** Dieser Abschnitt ist aktiviert, wenn mindestens ein Block für zufällige Effekte über das Elementdialogfeld "Zufällige Effekte" auf der Registerkarte "Felder & Effekte" erstellt wird und das Kontrollkästchen zum Anzeigen von Parametervorhersagen für diesen Block ausgewählt ist. Wenn mehrere solche Blöcke angegeben wurden, können Sie auswählen, dass die Ergebnisse in separaten Datasets oder Dateien gespeichert oder in einem einzelnen Dataset oder einer einzelnen Datei kombiniert werden sollen.

## Modellansicht

Die Prozedur erstellt ein Modellobjekt im Viewer. Wenn Sie dieses Objekt durch einen Doppelklick aktivieren, erhalten Sie eine interaktive Ansicht des Modells.

Standardmäßig wird die Ansicht "Modellzusammenfassung" angezeigt. Um eine andere Modellansicht anzuzeigen, wählen Sie sie aus den Piktogrammen aus.

Als Alternative zum Modellobjekt können Sie Pivot-Tabellen und Diagramme generieren, indem Sie **Pivot-Tabellen und Diagramme** in der Gruppe "Ausgabe anzeigen" auf der Registerkarte "Ausgabe" des Dialogfelds "Optionen" (Bearbeiten > Optionen) auswählen. In den folgenden Themen wird das Modellobjekt beschrieben.

## Modellübersicht

Diese Ansicht bietet eine momentane, übersichtliche Zusammenfassung des Modells und seiner Anpassungsgüte.

**Tabelle.** In der Tabelle sind das Ziel, die Wahrscheinlichkeitsverteilung und die Verknüpfungsfunktion zu sehen, die in den Zieleinstellungen angegeben wurden. Wenn das Ziel durch Ereignisse und Versuche definiert ist, wird die Zelle aufgeteilt und zeigt das Ereignisfeld und das Feld für die Versuche oder eine feste Anzahl von Versuchen an. Außerdem werden das Akaike-Informationskriterium mit Korrektur für endliche Stichproben (AICC) und das Bayes-Informationskriterium (BIC) angezeigt.

- *Akaike mit Korrektur.* Ein Maß für die Auswahl und den Vergleich gemischter Modelle auf der Basis der  $-2$  (eingeschränkten) Log-Likelihood. Kleinere Werte stehen für bessere Modelle. Das AICC "korrigiert" das AIC für kleine Stichprobenumfänge. Wenn der Stichprobenumfang zunimmt, konvergiert das AICC zu dem AIC.
- *Bayessch.* Ein Maß für die Auswahl und den Vergleich von Modellen auf der Basis der  $-2$  Log-Likelihood. Kleinere Werte stehen für bessere Modelle. BIC "penalisiert" überparametrisierte Modelle ebenfalls (beispielsweise Modelle mit einer großen Anzahl Eingaben), allerdings strikter als AIC.

**Diagramme.** Bei kategorialen Zielen zeigt ein Diagramm die Genauigkeit des endgültigen Modells an, also den Prozentsatz der korrekten Klassifizierungen.

## Datenstruktur

Diese Ansicht bietet einen Überblick über die von Ihnen angegebene Datenstruktur und hilft Ihnen bei der Überprüfung, ob die Subjekte und Messwiederholungen richtig angegeben wurde. Die beobachteten Informationen für das erste Subjekt werden für jedes Subjektfeld und Messwiederholungsfeld sowie für das Ziel angegeben. Außerdem wird die Anzahl der Stufen für die einzelnen Subjektfelder und Messwiederholungsfelder angezeigt.

## Vorhergesagt/Beobachtet

Für stetige Ziele, einschließlich Ziele, die als Ereignisse/Versuche angegeben sind, zeigt diese Ansicht ein klassiertes Streudiagramm, das die vorhergesagten Werte auf der vertikalen Achse in Abhängigkeit von den beobachteten Werten auf der horizontalen Achse darstellt. Idealerweise sollten die Werte entlang einer 45-Grad-Linie liegen. In dieser Ansicht können Sie erkennen, ob bestimmte Datensätze vom Modell besonders schlecht vorhergesagt werden.

## Klassifikation

Bei kategorialen Zielen wird hiermit die Kreuzklassifikation der beobachteten Werte in Abhängigkeit von den vorhergesagten Werten in einer Heat-Map angezeigt, zuzüglich des Gesamtprozentsatzes der korrekten Werte.

**Tabellenstile.** Es sind verschiedene Anzeigestile verfügbar, auf die über die Dropdown-Liste **Stil** zugegriffen werden kann.

- **Zeilenprozentwerte.** Zeigt die Zeilenprozentsätze (die Zellenhäufigkeiten, ausgedrückt als Prozentsatz der Gesamtzeilenanzahl) in den Zellen an. Dies ist der Standardwert.
- **Zellenhäufigkeit.** Hiermit werden die Zellenhäufigkeiten in den Zellen angezeigt. Die Schattierung für die Heat-Map beruht weiterhin auf den Zeilenprozentsätzen.
- **Heat-Map.** Zeigt in den Zellen keine Werte, sondern nur die Schattierung an.
- **Komprimiert.** Zeigt in den Zellen keine Zeilen- oder Spaltenüberschriften oder Werte an. Diese Option kann nützlich sein, wenn das Ziel sehr viele Kategorien aufweist.

**Fehlende Werte.** Wenn Datensätze fehlende Werte im Ziel aufweisen, werden diese unter allen gültigen Werten in der Zeile (**Fehlend**) angezeigt. Datensätze mit fehlenden Werten tragen nicht zum Wert von "Gesamtprozent korrekt" bei.

**Mehrere Ziele.** Bei mehreren kategorialen Zielen wird jedes Ziel in einer separaten Tabelle dargestellt. Dazu gibt es eine Dropdown-Liste **Ziel**, über die gesteuert werden kann, welches Ziel angezeigt wird.

**Große Tabellen.** Wenn das angezeigte Ziel mehr als 100 Kategorien aufweist, wird keine Tabelle angezeigt.

## Feste Effekte

Diese Ansicht zeigt die Größe der einzelnen festen Effekte im Modell.

**Stile.** Für die Diagramme sind verschiedene Anzeigestile verfügbar, auf die über die Dropdown-Liste **Stil** zugegriffen werden kann.

- **Diagramm.** In diesem Diagramm sind die Effekte von oben nach unten in der Reihenfolge sortiert, in der sie in den Einstellungen für "Feste Effekte" angegeben wurden. Verbindungslinien im Diagramm sind basierend auf der Effektsignifikanz gewichtet, wobei eine größere Linienbreite signifikanteren Effekten entspricht (kleinere  $p$ -Werte). Dies ist der Standardwert.
- **Tabelle.** Diese Ansicht zeigt eine ANOVA-Tabelle für das Gesamtmodell und die einzelnen Modelleffekte. Die einzelnen Effekte sind von oben nach unten in der Reihenfolge sortiert, in der sie in den Einstellungen unter "Feste Effekte" angegeben wurden.

**Signifikanz.** Mit dem Schieberegler "Signifikanz" kann gesteuert werden, welche Effekte in der Ansicht angezeigt werden. Effekte, deren Signifikanzwerte größer als der Wert des Schiebereglers sind, werden ausgeblendet. Dadurch wird das Modell nicht verändert, doch Sie können sich ganz problemlos auf die wichtigsten Effekte konzentrieren. Standardmäßig ist der Wert 1,00 eingestellt, sodass keine Effekte basierend auf der Signifikanz herausgefiltert werden.

## Feste Koeffizienten

Diese Ansicht zeigt den Wert der einzelnen festen Koeffizienten im Modell. Hinweis: Faktoren (kategoriale Prädiktoren) sind innerhalb des Modells indikatorcodiert, sodass Faktoren, die **Effekte** enthalten, in der Regel mehrere zugehörige **Koeffizienten** aufweisen. Mit Ausnahme der Kategorie für den redundanten Koeffizienten erhält jede Kategorie einen solchen Koeffizienten.

**Stile.** Es sind verschiedene Anzeigestile verfügbar, auf die über die Dropdown-Liste **Stil** zugegriffen werden kann.

- **Diagramm.** In diesem Diagramm wird zuerst der konstante Term angezeigt. Anschließend werden die Effekte von oben nach unten in der Reihenfolge sortiert, in der sie in den Einstellungen "Feste Effekte" angegeben wurden. In Faktoren, die Effekte enthalten, werden die Koeffizienten in aufsteigender Reihenfolge der Datenwerte sortiert. Verbindungslinien im Diagramm sind farbig dargestellt und basierend auf der Koeffizientensignifikanz gewichtet, wobei eine größere Linienbreite signifikanteren Koeffizienten entspricht (kleinere  $p$ -Werte). Dies ist der Standardstil.
- **Tabelle.** Diese Tabelle zeigt die Werte, Signifikanztests und Konfidenzintervalle für die einzelnen Modellkoeffizienten. Nach dem konstanten Term sind die einzelnen Effekte von oben nach unten in der Reihenfolge sortiert, in der sie in den Einstellungen für "Feste Effekte" angegeben wurden. In Faktoren, die Effekte enthalten, werden die Koeffizienten in aufsteigender Reihenfolge der Datenwerte sortiert.

**Multinomial.** Wenn die Multinomialverteilung in Kraft ist, steuert die Dropdown-Liste "Multinomial", welche Zielkategorie angezeigt werden soll. Die Sortierreihenfolge der Werte in der Liste richtet sich nach den Angaben unter "Erstellungsoptionen".

**Exponentialverteilung.** Hiermit werden exponentielle Koeffizientenschätzungen und Konfidenzintervalle für bestimmte Modelltypen angegeben, einschließlich "Binäre logistische Regression" (Binomialverteilung und Logit-Verknüpfung), "Nominale logistische Regression" (multinomiale Verteilung und Logit-Verknüpfung), "Negative binomiale Regression" (negative Binomialverteilung und Log-Verknüpfung) und Log-Linear-Modell (Poisson-Verteilung und Log-Verknüpfung).

**Signifikanz.** Mit dem Schieberegler "Signifikanz" wird gesteuert, welche Koeffizienten in der Ansicht angezeigt werden. Koeffizienten, deren Signifikanzwerte größer als der Wert des Schiebereglers sind, werden ausgeblendet. Dadurch wird das Modell nicht verändert, doch Sie können sich ganz problemlos auf die wichtigsten Koeffizienten konzentrieren. Standardmäßig ist der Wert 1,00 eingestellt, sodass keine Koeffizienten basierend auf der Signifikanz herausgefiltert werden.

## Kovarianzen der Zufallseffekte

Damit wird die Kovarianzmatrix der Zufallseffekte (**G**) angezeigt.

**Stile.** Es sind verschiedene Anzeigestile verfügbar, auf die über die Dropdown-Liste **Stil** zugegriffen werden kann.

- **Kovarianzwerte.** Dies ist eine Heat-Map der Kovarianzmatrix, in der die Effekte von oben nach unten in der Reihenfolge sortiert sind, in der sie in den Einstellungen für "Feste Effekte" angegeben wurden. Die Farben im Korrelogramm entsprechen den im Schlüssel angezeigten Zellenwerten. Dies ist der Standardwert.
- **Korrelogramm.** Dies ist eine Heat-Map der Kovarianzmatrix.
- **Komprimiert.** Dies ist eine Heat-Map der Kovarianzmatrix ohne Zeilen- und Spaltenüberschriften.

**Blöcke.** Wenn mehrere Blöcke für zufällige Effekte vorhanden sind, gibt es eine Dropdown-Liste mit dem Titel "Block", die zur Auswahl des anzuzeigenden Blocks dient.

**Gruppen.** Wenn ein Block für zufällige Effekte eine Gruppenspezifikation aufweist, gibt es eine Dropdown-Liste mit dem Titel "Gruppe", in der die anzuzeigende Stufe ausgewählt werden kann.

**Multinomial.** Wenn die Multinomialverteilung in Kraft ist, steuert die Dropdown-Liste "Multinomial", welche Zielkategorie angezeigt werden soll. Die Sortierreihenfolge der Werte in der Liste richtet sich nach den Angaben unter "Erstellungsoptionen".

## Kovarianzparameter

Diese Ansicht zeigt die Kovarianzparameterschätzungen und verwandte Statistiken für Residuum und Zufallseffekte. Dies sind erweiterte, aber dennoch grundlegende Ergebnisse, die Informationen darüber bieten, ob die Kovarianzstruktur geeignet ist.

**Zusammenfassende Tabelle.** Dies ist eine Kurzübersicht für die Anzahl der Parameter in den Kovarianzmatrizen für das Residuum (**R**) und den Zufallseffekt (**G**), den Rang (Anzahl der Spalten) in den Designmatrizen für den festen Effekt (**X**) und den Zufallseffekt (**Z**) und die Anzahl der Subjekte, die durch die Subjektfelder zur Definition der Datenstruktur festgelegt sind.

**Tabelle für Kovarianzparameter.** Für den ausgewählten Effekt werden Schätzung, Standardfehler und Konfidenzintervall für jeden Kovarianzparameter angezeigt. Die Anzahl der angezeigten Parameter hängt von der Kovarianzstruktur für den Effekt und, bei Blöcken für zufällige Effekte, von der Anzahl der Effekte im Block ab. Wenn Sie sehen, dass die außerhalb der Diagonalen liegenden Parameter nicht signifikant sind, können Sie eine einfachere Kovarianzstruktur verwenden.

**Effekte.** Wenn Blöcke für zufällige Effekte vorhanden sind, gibt es eine Dropdown-Liste mit dem Titel "Effekt", die zur Auswahl des anzuzeigenden Residuums oder des Blocks für zufällige Effekte dient. Der Residualeffekt steht immer zur Verfügung.

**Gruppen.** Wenn ein Residuum oder ein Block für zufällige Effekte eine Gruppenspezifikation aufweist, gibt es eine Dropdown-Liste mit dem Titel "Gruppe", in der die anzuzeigende Stufe ausgewählt werden kann.

**Multinomial.** Wenn die Multinomialverteilung in Kraft ist, steuert die Dropdown-Liste "Multinomial", welche Zielkategorie angezeigt werden soll. Die Sortierreihenfolge der Werte in der Liste richtet sich nach den Angaben unter "Erstellungsoptionen".

## Geschätzte Mittelwerte: Signifikante Effekte

Hierbei handelt es sich um Diagramme, die für die 10 "signifikantesten" festen Effekte für alle Faktoren angezeigt werden. Zuerst kommen die Dreiwegeinteraktionen, dann die Zweiwegeinteraktionen und schließlich die Haupteffekte. Das Diagramm zeigt den vom Modell geschätzten Zielwert auf der vertikalen Achse für jeden Wert des Haupteffekts (bzw. des ersten aufgeführten Effekts in einer Interaktion) auf der horizontalen Achse an. Für jeden Wert des zweiten in einer Interaktion aufgelisteten Effekts wird eine separate Linie erstellt. Für jeden Wert des dritten aufgelisteten Effekts in einer Dreiwegeinteraktion wird ein separates Diagramm erstellt. Alle anderen Prädiktoren werden konstant gehalten. Es gewährt eine nützliche Visualisierung der Effekte der einzelnen Prädiktorkoeffizienten auf dem Ziel. Beachten Sie: Wenn keine Prädiktoren signifikant sind, werden keine geschätzten Mittel produziert.

**Konfidenz.** Zeigt die oberen und die unteren Konfidenzgrenzen für die Randmittel an, wobei das im Rahmen der Erstellungsoptionen angegebene Konfidenzniveau verwendet wird.

## Geschätzte Mittelwerte: Benutzerdefinierte Effekte

Hierbei handelt es sich um Tabellen und Diagramme für vom Benutzer angeforderte feste Effekte für alle Faktoren.

**Stile.** Für die Diagramme sind verschiedene Anzeigestile verfügbar, auf die über die Dropdown-Liste **Stil** zugegriffen werden kann.

- **Diagramm.** Dieser Stil zeigt ein Liniendiagramm des vom Modell geschätzten Zielwerts auf der vertikalen Achse für jeden Wert des Haupteffekts (bzw. des ersten aufgeführten Effekts in einer Interaktion) auf der horizontalen Achse an. Für jeden Wert des zweiten in einer Interaktion aufgelisteten Effekts wird eine separate Linie erstellt. Für jeden Wert des dritten aufgelisteten Effekts in einer Dreiwegeinteraktion wird ein separates Diagramm erstellt. Alle anderen Prädiktoren werden konstant gehalten.

Wenn Kontraste angefordert werden, wird ein weiteres Diagramm erstellt, um die Stufen für das Kontrastfeld zu vergleichen. Bei Interaktionen wird ein Diagramm für jede Stufenkombination der Effekte (abgesehen vom Kontrastfeld) angezeigt. Bei **paarweisen** Kontrasten handelt es sich um ein Abstandsnetzdiagramm, also um eine grafische Darstellung der Vergleichstabelle, in der die Abstände zwischen Knoten im Netz den Unterschieden zwischen Stichproben entsprechen. Gelbe Linien entsprechen statistisch signifikanten Unterschieden, schwarze Linien nicht signifikanten Unterschieden. Wenn Sie die Maus über eine Linie im Netz bewegen, wird eine QuickInfo mit der angepassten Signifikanz des Unterschieds zwischen den durch die Linie verbundenen Knoten angezeigt.

Bei **Abweichungs**kontrasten wird ein Balkendiagramm mit dem vom Modell geschätzten Zielwert der vertikalen Achse und den Werten des Kontrastfelds auf der horizontalen Achse angezeigt. Bei Interaktionen wird ein Diagramm für jede Stufenkombination der Effekte (abgesehen vom Kontrastfeld) angezeigt. Die Balken zeigen die Differenz zwischen den einzelnen Stufen des Kontrastfelds und dem Gesamtmittelwert, der durch eine schwarze horizontale Linie dargestellt wird.

Bei **einfachen** Kontrasten, wird ein Balkendiagramm mit dem vom Modell geschätzten Zielwert der vertikalen Achse und den Werten des Kontrastfelds auf der horizontalen Achse angezeigt. Bei Interaktionen wird ein Diagramm für jede Stufenkombination der Effekte (abgesehen vom Kontrastfeld) angezeigt. Die Balken zeigen die Differenz zwischen den einzelnen Stufen des Kontrastfelds (abgesehen von der letzten) und der letzten Stufe, die durch eine schwarze horizontale Linie dargestellt wird.

- **Tabelle.** Dieser Stil zeigt eine Tabelle mit dem vom Modell geschätzten Zielwert, seinem Standardfehler und dem Konfidenzintervall für jede Stufenkombination der Felder im Effekt an. Alle anderen Prädiktoren werden konstant gehalten.

Wenn Kontraste angefordert wurden, wird eine weitere Tabelle mit Schätzung, Standardfehler, Signifikanztest und Konfidenzintervall für jeden Kontrast angezeigt. Bei Interaktionen wird eine separate Gruppe von Zeilen für jede Stufenkombination der Effekte (abgesehen vom Kontrastfeld) angezeigt. Außerdem wird eine Tabelle mit den Gesamttestergebnissen angezeigt. Bei Interaktionen gibt es einen separaten Gesamttest für jede Stufenkombination der Effekte (abgesehen vom Kontrastfeld).

**Konfidenz.** Schaltet zwischen der Anzeige der oberen und der unteren Konfidenzgrenzen für die Randmittel um, wobei das im Rahmen der Erstellungsoptionen angegebene Konfidenzniveau verwendet wird.

**Layout.** Schaltet zwischen den Layouts für das Diagramm der paarweisen Kontraste um. Das Kreislayout zeigt weniger Kontraste als das Netzlayout, vermeidet jedoch das Überlappen von Linien.

## Modellauswahl für loglineare Analyse

---

Mit der Prozedur "Modellauswahl für loglineare Analyse" werden Mehrwegekreuztabellen (Kontingenztabellen) analysiert. Sie können damit hierarchische loglineare Modelle mehrdimensionalen Kreuztabellen anpassen. Dabei wird ein iterativer, sich proportional anpassender Algorithmus verwendet. Diese Prozedur hilft Ihnen, herauszufinden, welche kategorialen Variablen verbunden sind. Um Modelle zu erstellen, stehen als Methode der erzwungene Einschluss sowie Rückwärtselimination zur Verfügung. Bei gesättigten Modellen können Sie Parameterschätzungen und Tests auf partiellen Zusammenhang anfordern. Bei einem gesättigten Modell wird der Wert aller Zellen um 0,5 erhöht.

**Beispiel.** Bei einer Studie über das Verbraucherverhalten hinsichtlich zweier Waschmittelmarken zählten Forscher die Personen jeder Gruppe und kombinierten verschiedene Kategorien der Wasserhärte (weich, mittel oder hart), der vorausgegangenen Verwendung einer der Marken und der Waschtemperatur (30 oder 60 Grad). Es zeigte sich ein Zusammenhang zwischen Temperatur und Wasserhärte sowie bevorzugter Marke.

**Statistik** Häufigkeiten, Residuen, Parameterschätzungen, Standardfehler, Konfidenzintervalle, Tests auf partiellen Zusammenhang. Bei benutzerdefinierten Modellen stehen Residuen- und Normalverteilungsdiagramme zur Verfügung.

Erläuterungen der Daten zur Modellauswahl für loglineare Analyse

**Daten.** Die Faktorvariablen sind kategorial. Alle zu analysierenden Variablen müssen numerisch sein. Sie können kategoriale Zeichenfolgevariablen in numerische Variablen umcodieren, bevor Sie mit der Analyse zur Modellauswahl beginnen.

Vermeiden Sie, zahlreiche Variablen mit einer Vielzahl von Ausprägungen zu definieren. Solche Angaben können zu zahlreichen Zellen mit wenigen Beobachtungen führen, sodass die resultierenden Chi-Quadrat-Werte unbrauchbar werden.

**Verwandte Prozeduren.** Mit der Prozedur "Modellauswahl" können Sie die in einem Modell benötigten Terme leichter finden. Anschließend können Sie das Modell unter Verwendung der Allgemeinen loglinearen Analyse oder der Logit-loglinearen Analyse weitergehend auswerten. Sie können "Automatisch umcodieren" verwenden, um Zeichenfolgevariablen umzucodieren. Wenn eine numerische Variable über leere Kategorien verfügt, verwenden Sie "Umcodieren", um fortlaufende ganzzahlige Werte zu erhalten.

Anfordern einer Modellauswahl für loglineare Analyse

Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Loglinear > Modellauswahl...**

1. Wählen Sie mindestens zwei numerische kategoriale Faktoren aus.
2. Wählen Sie aus der Faktorenliste mindestens eine Faktorvariable aus und klicken Sie auf **Bereich definieren**.
3. Definieren Sie den Wertebereich für jede Faktorvariable.
4. Wählen Sie im Gruppenfeld "Modell erstellen" eine Option aus.

Wahlweise können Sie eine Zellengewichtungsvariable auswählen, um strukturelle Nullen anzugeben.

## Loglineare Analyse: Bereich definieren

Sie müssen den Bereich der Kategorien für jede Faktorvariable angeben. Die Werte in den Feldern "Minimum" und "Maximum" entsprechen der niedrigsten bzw. höchsten Kategorie der Faktorvariablen. Beide Werte müssen ganzzahlig sein. Der Minimalwert muss kleiner sein als der Maximalwert. Wenn Sie beispielsweise als Minimalwert 1 und als Maximalwert 3 angeben, werden nur die Werte 1, 2 und 3 verwendet. Wiederholen Sie diesen Vorgang für jede Faktorvariable.

## Loglineare Analyse: Modell

**Modell angeben.** Ein gesättigtes Modell enthält alle Haupteffekte der Faktoren sowie alle Interaktionen zwischen den einzelnen Faktoren. Wählen Sie **Anpassen** aus, um die modellbildende Klasse für ein ungesättigtes Modell anzugeben.

**Modellbildende Klasse.** Eine modellbildende Klasse ist eine Liste der ranghöchsten Terme, in der Faktoren auftreten. Ein hierarchisches Modell enthält die Terme, die die modellbildende Klasse definieren, und alle verwandten Terme niedrigerer Ordnung. Angenommen, Sie wählen aus der Faktorenliste die Variablen  $A$ ,  $B$  und  $C$  und aus dem Dropdown-Listefeld "Term(e) erstellen" die Option **Interaktion** aus. Das resultierende Modell enthält die angegebene 3-Wege-Interaktion  $A*B*C$ , die 2-Wege-Interaktionen  $A*B$ ,  $A*C$  und  $B*C$  sowie die Haupteffekte für  $A$ ,  $B$  und  $C$ . Geben Sie nicht die Verwandten der unteren Ordnung in der Generierungsklasse an.

Für die ausgewählten Faktoren:

### Interaktion

Hiermit wird der Interaktionsterm mit der höchsten Ordnung von allen ausgewählten Variablen erstellt. Dies ist der Standardwert.

### Haupteffekte

Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

### Alle 2-Wege

Hiermit werden alle möglichen Zweibegeinteraktionen der ausgewählten Variablen erstellt.

### Alle 3-Wege

Hiermit werden alle möglichen Dreibegeinteraktionen der ausgewählten Variablen erzeugt.

### Alle 4-Wege

Erstellt alle möglichen Vierbegeinteraktionen der ausgewählten Variablen.

## Alle 5-Wege

Hiermit werden alle möglichen Fünfwegeinteraktionen der ausgewählten Variablen erzeugt.

## Erstellen von Termen und benutzerdefinierten Termen

### Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie nicht verschachtelte Terme eines bestimmten Typs (beispielsweise Haupteffekte) für alle Kombinationen eines ausgewählten Sets von Faktoren und Kovariaten einschließen wollen.

### Benutzerdefinierte Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie verschachtelte Terme einschließen oder einen Term explizit Variable für Variable erstellen wollen. Die Erstellung eines verschachtelten Terms besteht aus folgenden Schritten:

## Modellauswahl für loglineare Analyse: Optionen

**Anzeigen.** Sie können **Häufigkeiten**, **Residuen** oder beides auswählen. In einem gesättigten Modell sind beobachtete und erwartete Häufigkeiten gleich. Die Residuen sind gleich 0.

**Diagramm.** Für benutzerdefinierte Modelle können Sie die folgenden Diagrammtypen auswählen: **Residuen** und/oder **Normalverteilung**. Hiermit können Sie ermitteln, wie gut sich ein Modell auf die Daten passt.

**Für gesättigtes Modell anzeigen.** Sie können für ein gesättigtes Modell **Parameterschätzungen** auswählen. Mithilfe von Parameterschätzungen können Sie die im Modell vernachlässigbaren Terme festlegen. Ebenfalls verfügbar ist eine Assoziationstabelle mit Tests auf partiellen Zusammenhang. Bei Tabellen mit vielen Faktoren ist diese Option rechenintensiv.

**Modellkriterien.** Zur Anforderung von Parameterschätzungen wird ein iterativer, sich proportional anpassender Algorithmus verwendet. Ein oder mehrere Schätzkriterien können durch die Angabe der folgenden Optionen überschrieben werden: **Anzahl der Iterationen**, **Konvergenz** oder **Delta** (ein Wert, der zu allen Zellenhäufigkeiten in gesättigten Modellen addiert wird).

## Zusätzliche Funktionen beim Befehl HILOGLINEAR

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl **CWEIGHT** können Sie Zellengewichtungen in Matrixform angeben.
- Mit dem Unterbefehl **DESIGN** können Sie die Analysen mehrerer Modelle mit einem einzigen Befehl starten.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Allgemeine loglineare Analyse

---

Mit der Prozedur "Allgemeine loglineare Analyse" werden die Häufigkeiten der Beobachtungen analysiert, die in einer Kreuztabelle oder einer Kontingenztafel jeder Kategorie der Kreuzklassifikation zugeordnet sind. Jede Kreuzklassifikation in der Tabelle bildet eine Zelle und jede kategoriale Variable wird als Faktor bezeichnet. Die abhängige Variable stellt die Anzahl der Fälle (Häufigkeit) in einer Zelle der Kreuztabelle dar und die erklärenden Variablen sind die Faktoren und Kovariaten. Mit dieser Prozedur werden Maximum-Likelihood-Parameter von hierarchischen und nicht hierarchischen loglinearen Modellen unter Verwendung der Newton-Raphson-Methode geschätzt. Es können entweder eine Poisson-Verteilung oder eine multinomiale Verteilung analysiert werden.

Sie können bis zu 10 Faktoren auswählen, um die Zellen einer Tabelle zu definieren. Mit einer Zellenstrukturvariablen können strukturelle Nullen für unvollständige Tabellen definiert, ein Offset-Term in das Modell einbezogen, ein Log-Rate-Modell angepasst oder die Methode der Anpassung an Tabellen mit gegebenen Randwerten angewendet werden. Mit Kontrastvariablen können verallgemeinerte Log-Odds-Verhältnisse (GLOR) berechnet werden.

Informationen zum Modell und Statistiken der Anpassungsgüte werden automatisch angezeigt. Sie können sich auch eine Vielzahl von Statistiken und Diagrammen anzeigen lassen oder Residuen und vorhergesagte Werte im aktiven Dataset speichern.

**Beispiel.** Mit Daten aus einem Bericht über Verkehrsunfälle in Florida wird der Zusammenhang zwischen angelegtem Sicherheitsgurt und einem tödlichen oder nicht tödlichen Ausgang der Verletzungen bestimmt. Dem Odds-Verhältnis zufolge ergibt sich ein signifikanter Hinweis auf einen Zusammenhang.

**Statistiken.** Beobachtete und erwartete Häufigkeiten; unbearbeitete, angepasste und Abweichungsresiduen; Designmatrix; Parameterschätzungen; Odds-Verhältnis; Log-Odds-Verhältnis; GLOR; Wald-Statistik; Konfidenzintervalle. Diagramme: angepasste Residuen, abweichende Residuen und Normalverteilungsdiagramme.

Erläuterungen der Daten für die allgemeine loglineare Analyse

**Daten.** Faktoren sind kategorial und Zellenkovariaten sind stetig. Wenn das Modell eine Kovariate enthält, wird den Zellen jeweils das Kovariatenmittel für Fälle in der betreffenden Zelle zugewiesen. Kontrastvariablen sind stetig. Sie dienen der Berechnung der verallgemeinerten Log-Odds-Verhältnisse. Die Werte der Kontrastvariablen sind die Koeffizienten für die lineare Kombination der Logarithmen der erwarteten Zellenhäufigkeiten.

Mit einer Zellenstrukturvariablen werden Gewichtungen zugewiesen. Wenn zum Beispiel einige der Zellen strukturelle Nullen sind, hat die Zellenstrukturvariable einen Wert von 0 oder 1. Verwenden Sie keine Zellenstrukturvariablen, um zusammengefasste Daten zu gewichten. Wählen Sie stattdessen aus dem Menü "Daten" den Befehl **Fälle gewichten** aus.

**Annahmen.** In der allgemeinen loglinearen Analyse sind zwei Verteilungen verfügbar: Poisson-Verteilung und multinomiale Verteilung.

Unter der Annahme einer Poisson-Verteilung:

- Vor der Untersuchung wird die gesamte Stichprobengröße nicht festgesetzt oder die Analyse ist für die gesamte Stichprobengröße nicht bedingt.
- Das Vorkommen einer Beobachtung in einer Zelle ist statistisch unabhängig von den Zellenhäufigkeiten anderer Zellen.

Unter der Annahme einer multinomialen Verteilung:

- Die gesamte Stichprobengröße wird festgesetzt oder die Analyse ist für die gesamte Stichprobengröße bedingt.
- Die Zellenhäufigkeiten sind nicht statistisch unabhängig.

**Verwandte Prozeduren.** Verwenden Sie die Prozedur "Kreuztabellen", um Kreuztabellen zu untersuchen. Verwenden Sie die Prozedur "Logit-Loglinear", wenn eine oder mehrere kategoriale Variablen als Antwortvariablen und die anderen als erklärende Variablen betrachtet werden können.

Berechnen einer allgemeinen loglinearen Analyse

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Loglinear > Allgemein...**

2. Wählen Sie im Dialogfeld "Allgemeine loglineare Analyse" bis zu 10 Faktorvariablen aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie Zellenkovariaten aus.
- Wählen Sie eine Zellenstrukturvariable aus, um strukturelle Nullen zu definieren, oder einen Offset-Term einzubeziehen.
- Wählen Sie eine Kontrastvariable aus.

## Allgemeine loglineare Analyse: Modell

**Modell angeben.** Ein gesättigtes Modell enthält alle Haupteffekte und Interaktionen mit Faktorvariablen. Es sind keine Terme mit Kovariaten enthalten. Wählen Sie **Anpassen** aus, um nur ein Subset von Interaktionen oder Interaktionen zwischen Faktoren und Kovariaten festzulegen.

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

**Terme im Modell.** Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von **Anpassen** können Sie die Haupteffekte und Interaktionen auswählen, die für Ihre Analyse von Interesse sind. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

Für die ausgewählten Faktoren und Kovariaten:

### Interaktion

Hiermit wird der Interaktionsterm mit der höchsten Ordnung von allen ausgewählten Variablen erstellt. Dies ist der Standardwert.

### Haupteffekte

Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

### Alle 2-Wege

Hiermit werden alle möglichen Zweiwegeinteraktionen der ausgewählten Variablen erstellt.

### Alle 3-Wege

Hiermit werden alle möglichen Dreiwegeinteraktionen der ausgewählten Variablen erzeugt.

### Alle 4-Wege

Erstellt alle möglichen Vierwegeinteraktionen der ausgewählten Variablen.

### Alle 5-Wege

Hiermit werden alle möglichen Fünfwegeinteraktionen der ausgewählten Variablen erzeugt.

## Erstellen von Termen und benutzerdefinierten Termen

### Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie nicht verschachtelte Terme eines bestimmten Typs (beispielsweise Haupteffekte) für alle Kombinationen eines ausgewählten Sets von Faktoren und Kovariaten einschließen wollen.

### Benutzerdefinierte Terme erstellen

Verwenden Sie diese Auswahl, wenn Sie verschachtelte Terme einschließen oder einen Term explizit Variable für Variable erstellen wollen. Die Erstellung eines verschachtelten Terms besteht aus folgenden Schritten:

## Allgemeine loglineare Analyse: Optionen

Mit der Prozedur "Allgemeine loglineare Analyse" werden Informationen zum Modell und Statistiken für Anpassungsgüte angezeigt. Zusätzlich können Sie eine oder mehrere der folgenden Optionen wählen:

**Anzeigen.** Für die Anzeige sind verschiedene Statistiken verfügbar - beobachtete und erwartete Zellenhäufigkeit; unbearbeitete, korrigierte und Abweichungsresiduen; eine Designmatrix des Modells sowie Parameterschätzungen für das Modell.

**Diagramm.** Als Diagramme (nur für benutzerdefinierte Modelle) sind unter anderem zwei Streudiagrammmatrizen verfügbar (angepasste Residuen oder Abweichungsresiduen über beobachtete und erwartete Zellenhäufigkeiten). Sie können sich auch Normalverteilungsdiagramme und zentrierte Normalverteilungsdiagramme von angepassten Residuen oder Abweichungsresiduen anzeigen lassen.

**Konfidenzintervall.** Das Konfidenzintervall für die Parameterschätzungen kann angepasst werden.

**Kriterien.** Die Newton-Raphson-Methode wird für Maximum-Likelihood-Parameterschätzungen verwendet. Sie können neue Werte für die Höchstzahl der Iterationen sowie das Konvergenzkriterium und Delta (eine für Anfangsapproximationen zu allen Zellen addierte Konstante) eingeben. Delta verbleibt in den Zellen für gesättigte Modelle.

## Allgemeine loglineare Analyse: Speichern

Wählen Sie die Werte aus, die Sie als neue Variablen im aktiven Dataset speichern möchten. Um jede gespeicherte Variable mit einem eindeutigen Namen zu versehen, wird das Suffix  $n$  schrittweise geändert.

Die gespeicherten Werte beziehen sich auf die zusammengefassten Daten (Zellen in der Kontingenztabelle), auch wenn die Daten im Dateneditor nach einzelnen Beobachtungen aufgezeichnet sind. Wenn Sie Residuen oder Vorhersagewerte für nicht aggregierte Daten speichern, wird der gespeicherte Wert einer Zelle in der Kontingenztabelle für jeden in dieser Zelle vorhandenen Fall in den Dateneditor aufgenommen. Um die gespeicherten Werte sinnvoll zu nutzen, sollten Sie die Daten aggregieren, um Zellenhäufigkeiten zu erhalten.

Es können vier Typen von Residuen gespeichert werden: unbearbeitete, standardisierte und angepasste Residuen sowie Abweichungsresiduen. Außerdem können die vorhergesagten Werte gespeichert werden.

- *Residuen*. Auch bekannt als einfaches Residuum oder Rohresiduum. Die Differenz zwischen der beobachteten und der erwarteten Zellenanzahl.
- *Standardisierte Residuen*. Das Residuum dividiert durch eine Schätzung seines Standardfehlers. Standardisierte Residuen werden auch als Pearson-Residuen bezeichnet.
- *Angepasste Residuen*. Das standardisierte Residuum dividiert durch den geschätzten Standardfehler. Wenn das ausgewählte Modell zutrifft, sind die angepassten Residuen asymptotisch standardnormalverteilt und daher für den Test auf Normalverteilung den standardisierten Residuen vorzuziehen.
- *Abweichungsresiduen*. Die Quadratwurzel mit Vorzeichen eines einzelnen Beitrags zur Chi-Quadrat-Statistik für Likelihood-Quotienten (G-Quadrat), wobei das Vorzeichen das Vorzeichen des Residuums (beobachtete Anzahl minus erwartete Anzahl) ist. Abweichungsresiduen sind asymptotisch standardnormalverteilt.

## Zusätzliche Funktionen beim Befehl GENLOG

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl GERESID können lineare Kombinationen von beobachteten und erwarteten Zellenhäufigkeiten berechnet und Residuen, standardisierte Residuen und korrigierte Residuen der jeweiligen Kombination ausgegeben werden.
- Mit dem Unterbefehl CRITERIA kann der voreingestellte Schwellenwert zur Redundanzprüfung geändert werden.
- Mit dem Unterbefehl PRINT können die standardisierten Residuen angezeigt werden.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Logit-loglineare Analyse

Die Prozedur "Logit-loglineare Analyse" analysiert die Beziehung zwischen abhängigen Variablen (Antworten) und unabhängigen Variablen (erklärenden Variablen). Die abhängigen Variablen sind immer kategorial, während die unabhängigen Variablen kategorial (Faktoren) sein können. Andere unabhängige Variablen (Zellenkovariaten) können stetig sein, werden aber nicht auf einer fallweisen Basis zugewiesen. Das gewichtete Kovariatenmittel einer Zelle wird dieser Zelle zugewiesen. Der Logarithmus der Odds der abhängigen Variablen wird als die lineare Kombination von Parametern ausgedrückt. Eine multinomiale Verteilung wird automatisch angenommen; diese Modelle werden auch multinomiale Logit-Modelle genannt. Diese Prozedur schätzt die Parameter der Logit-loglinearen Modelle mithilfe des Algorithmus nach Newton-Raphson.

Sie können aus 1 bis 10 abhängigen und kombinierten Faktorvariablen auswählen. Mit einer Zellenstrukturvariablen können strukturelle Nullen für unvollständige Tabellen definiert, ein Offset-Term in das Modell einbezogen, ein Log-Rate-Modell angepasst oder die Methode der Anpassung an Tabellen mit gegebenen Randwerten angewendet werden. Mit Kontrastvariablen können verallgemeinerte Log-Odds-Verhältnisse (GLOR) berechnet werden. Die Werte der Kontrastvariablen sind die Koeffizienten für die lineare Kombination der Logarithmen der erwarteten Zellenhäufigkeiten.

Informationen zum Modell und Statistiken der Anpassungsgüte werden automatisch angezeigt. Sie können sich auch eine Vielzahl von Statistiken und Diagrammen anzeigen lassen oder Residuen und vorhergesagte Werte im aktiven Dataset speichern.

**Beispiel.** In einer Studie in Florida wurden 219 Alligatoren beobachtet. Wie hängt die Art der Alligatornahrung von deren Größe und den vier Seen ab, in denen sie leben? Die Studie ergab, dass die Quote (Odds) der Alligatoren, die statt Fisch Reptilien bevorzugen, bei kleineren Alligatoren 0,7 mal niedriger als bei größeren ist. Außerdem war die Quote der vorzugsweisen Ernährung durch Reptilien anstelle von Fisch im See Nr. 3 am höchsten.

**Statistik** Beobachtete und erwartete Häufigkeiten; unbearbeitete, angepasste und abweichende Residuen; Designmatrix; Parameterschätzungen; verallgemeinertes Log-Odds-Verhältnis; Wald-Statistik; Konfidenzintervalle. Diagramme: angepasste Residuen, abweichende Residuen und Normalverteilungsdiagramme.

Erläuterungen der Daten für die Logit-loglineare Analyse

**Daten.** Die abhängigen Variablen sind kategorial. Faktoren sind kategorial. Zellenkovariaten können stetig sein, aber wenn eine Kovariate im Modell enthalten ist, wird der mittlere Kovariatenwert für Fälle einer Zelle dieser Zelle zugewiesen. Kontrastvariablen sind stetig. Sie werden verwendet, um das verallgemeinerte Log-Odds-Verhältnis (GLOR) zu berechnen. Die Werte der Kontrastvariablen sind die Koeffizienten für die lineare Kombination der Logarithmen der erwarteten Zellenhäufigkeiten.

Mit einer Zellenstrukturvariablen werden Gewichtungen zugewiesen. Wenn zum Beispiel einige der Zellen strukturelle Nullen sind, hat die Zellenstrukturvariable einen Wert von 0 oder 1. Verwenden Sie keine Zellenstrukturvariablen, um zusammengefasste Daten zu gewichten. Verwenden Sie stattdessen den Befehl "Fälle gewichten" aus dem Menü "Daten".

**Annahmen.** Bei den Häufigkeiten innerhalb jeder Kombination der Kategorien von erklärenden Variablen wird eine multinomiale Verteilung angenommen. Unter der Annahme einer multinomialen Verteilung:

- Die gesamte Stichprobengröße wird festgesetzt oder die Analyse ist für die gesamte Stichprobengröße bedingt.
- Die Zellenhäufigkeiten sind nicht statistisch unabhängig.

**Verwandte Prozeduren.** Verwenden Sie die Prozedur "Kreuztabellen", um die Kontingenztabellen anzuzeigen. Verwenden Sie die Prozedur "Allgemeine loglineare Analyse", wenn Sie die Beziehung zwischen einer beobachteten Anzahl und einem Set von erklärenden Variablen analysieren möchten.

Berechnen einer Logit-loglinearen Analyse

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Loglinear > Logit...**

2. Wählen Sie im Dialogfeld "Logit-loglineare Analyse" mindestens eine abhängige Variable aus.
3. Wählen Sie mindestens eine Faktorvariable aus.

Die Gesamtzahl der abhängigen Variablen und der Faktorvariablen muss kleiner-gleich 10 sein.

Die folgenden Optionen sind verfügbar:

- Wählen Sie Zellenkovariaten aus.
- Wählen Sie eine Zellenstrukturvariable aus, um strukturelle Nullen zu definieren, oder einen Offset-Term einzubeziehen.
- Wählen Sie mindestens eine Kontrastvariable aus.

## Logit-loglineare Analyse: Modell

**Modell angeben.** Ein gesättigtes Modell enthält alle Haupteffekte und Interaktionen mit Faktorvariablen. Es sind keine Terme mit Kovariaten enthalten. Wählen Sie **Anpassen** aus, um nur ein Subset von Interaktionen oder Interaktionen zwischen Faktoren und Kovariaten festzulegen.

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

**Terme im Modell.** Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von **Anpassen** können Sie die Haupteffekte und Interaktionen auswählen, die für Ihre Analyse von Interesse sind. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

Für die ausgewählten Faktoren und Kovariaten:

#### **Interaktion**

Hiermit wird der Interaktionsterm mit der höchsten Ordnung von allen ausgewählten Variablen erstellt. Dies ist der Standardwert.

#### **Haupteffekte**

Hiermit wird ein Haupteffektterm für jede ausgewählte Variable erzeugt.

#### **Alle 2-Wege**

Hiermit werden alle möglichen Zweiwegeinteraktionen der ausgewählten Variablen erstellt.

#### **Alle 3-Wege**

Hiermit werden alle möglichen Dreiwegeinteraktionen der ausgewählten Variablen erzeugt.

#### **Alle 4-Wege**

Erstellt alle möglichen Vierwegeinteraktionen der ausgewählten Variablen.

#### **Alle 5-Wege**

Hiermit werden alle möglichen Fünfwegeinteraktionen der ausgewählten Variablen erzeugt.

Terme werden zum Design hinzugefügt, indem alle möglichen Kombinationen der abhängigen Terme mit jedem Term der Modellliste verbunden wird. Wenn die Option **Konstante für abhängige Variable einschließen** ausgewählt wird, so wird der Modellliste ein Einheitenterm (1) hinzugefügt.

Angenommen, die Variablen  $D1$  und  $D2$  sind die abhängigen Variablen. Die Prozedur "Logit-loglineare Analyse" erstellt eine Liste der abhängigen Terme ( $D1$ ,  $D2$ ,  $D1*D2$ ). Wenn die Liste "Terme im Modell"  $M1$  und  $M2$  enthält und eine Konstante eingeschlossen ist, enthält die Modellliste 1,  $M1$  und  $M2$ . Das resultierende Design enthält die Kombinationen aller Modellterme mit allen abhängigen Termen:

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

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

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

**Konstante für abhängige Variable einschließen.** Schließt eine Konstante für eine abhängige Variable in ein benutzerdefiniertes Modell ein.

## **Erstellen von Termen und benutzerdefinierten Termen**

### **Terme erstellen**

Verwenden Sie diese Auswahl, wenn Sie nicht verschachtelte Terme eines bestimmten Typs (beispielsweise Haupteffekte) für alle Kombinationen eines ausgewählten Sets von Faktoren und Kovariaten einschließen wollen.

### **Benutzerdefinierte Terme erstellen**

Verwenden Sie diese Auswahl, wenn Sie verschachtelte Terme einschließen oder einen Term explizit Variable für Variable erstellen wollen. Die Erstellung eines verschachtelten Terms besteht aus folgenden Schritten:

## **Logit-loglineare Analyse: Optionen**

Die Prozedur "Logit-loglineare Analyse" gibt Modellinformation und Statistiken über die Anpassungsgüte aus. Zusätzlich können Sie eine oder mehrere der folgenden Optionen wählen:

**Anzeigen.** Für die Anzeige sind verschiedene Statistiken verfügbar: beobachtete und erwartete Zellenhäufigkeit; unbearbeitete, angepasste und Abweichungsresiduen; eine Designmatrix des Modells sowie Parameterschätzungen für das Modell.

**Diagramm.** Für benutzerdefinierte Modelle sind zwei Streudiagrammmatrizen verfügbar (angepasste Residuen oder Abweichungsresiduen über beobachtete und erwartete Zellenhäufigkeiten). Sie können

auch Normalverteilungsdiagramme und trendbereinigte Normalverteilungsdiagramme von angepassten Residuen oder Abweichungsresiduen anzeigen.

**Konfidenzintervall.** Das Konfidenzintervall für die Parameterschätzungen kann angepasst werden.

**Kriterien.** Die Newton-Raphson-Methode wird für Maximum-Likelihood-Parameterschätzungen verwendet. Sie können neue Werte für die maximale Anzahl von Iterationen, das Konvergenzkriterium und einen Deltawert (eine Konstante, die für anfängliche Approximationen zu allen Zellen hinzugefügt wird) angeben. Delta verbleibt in den Zellen für gesättigte Modelle.

## Logit-loglineare Analyse: Speichern

Wählen Sie die Werte aus, die Sie als neue Variablen im aktiven Dataset speichern möchten. Um jede gespeicherte Variable mit einem eindeutigen Namen zu versehen, wird das Suffix  $n$  schrittweise geändert.

Die gespeicherten Werte beziehen sich auf die aggregierten Daten (auf Zellen in der Kontingenztabelle), auch wenn die Daten im Dateneditor als einzelne Beobachtungen aufgezeichnet wurden. Wenn Sie Residuen oder Vorhersagewerte für nicht aggregierte Daten speichern, wird der gespeicherte Wert einer Zelle in der Kontingenztabelle für jeden in dieser Zelle vorhandenen Fall in den Dateneditor aufgenommen. Um die gespeicherten Werte sinnvoll zu nutzen, sollten Sie die Daten aggregieren, um Zellenhäufigkeiten zu erhalten.

Es können vier Typen von Residuen gespeichert werden: unbearbeitete, standardisierte und angepasste Residuen sowie Abweichungsresiduen. Außerdem können die vorhergesagten Werte gespeichert werden.

- *Residuen.* Auch bekannt als einfaches Residuum oder Rohresiduum. Die Differenz zwischen der beobachteten und der erwarteten Zellenanzahl.
- *Standardisierte Residuen.* Das Residuum dividiert durch eine Schätzung seines Standardfehlers. Standardisierte Residuen werden auch als Pearson-Residuen bezeichnet.
- *Angepasste Residuen.* Das standardisierte Residuum dividiert durch den geschätzten Standardfehler. Wenn das ausgewählte Modell zutrifft, sind die angepassten Residuen asymptotisch standardnormalverteilt und daher für den Test auf Normalverteilung den standardisierten Residuen vorzuziehen.
- *Abweichungsresiduen.* Die Quadratwurzel mit Vorzeichen eines einzelnen Beitrags zur Chi-Quadrat-Statistik für Likelihood-Quotienten (G-Quadrat), wobei das Vorzeichen das Vorzeichen des Residuums (beobachtete Anzahl minus erwartete Anzahl) ist. Abweichungsresiduen sind asymptotisch standardnormalverteilt.

## Zusätzliche Funktionen beim Befehl GENLOG

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl GERESID können lineare Kombinationen von beobachteten und erwarteten Zellenhäufigkeiten berechnet und Residuen, standardisierte Residuen und korrigierte Residuen der jeweiligen Kombination ausgegeben werden.
- Mit dem Unterbefehl CRITERIA kann der voreingestellte Schwellenwert zur Redundanzprüfung geändert werden.
- Mit dem Unterbefehl PRINT können die standardisierten Residuen angezeigt werden.

See the *Command Syntax Reference* for complete syntax information.

## Sterbetafel

Es kann viele Situationen geben, in denen Sie die Verteilung der Zeiten zwischen zwei Ereignissen untersuchen möchten, wie z. B. die Dauer einer Beschäftigung (Zeit zwischen Einstellung und Kündigung eines Arbeitsplatzes). Diese Art von Daten umfasst in der Regel allerdings auch Fälle, für die das zweite Ereignis nicht aufgezeichnet ist (z. B. weil Arbeitnehmer am Ende der Studie immer noch in dem Unternehmen beschäftigt sind). Dafür kann es unterschiedliche Gründe geben: In einigen Fällen tritt das Ereignis einfach vor Abschluss der Studie nicht ein; in anderen Fällen kann der Status nicht bis zum Abschluss der Studie verfolgt werden; andere Fälle wiederum können aus Gründen, die von der Studie unabhängig

sind, nicht weiter weiterverfolgt werden (wie z. B. Beurlaubung oder Erkrankung eines Mitarbeiters). Zusammengefasst werden solche Fälle als **zensierte Fälle** bezeichnet. Sie verhindern die Anwendung der traditionellen Methoden wie des  $t$ -Tests oder der Regression auf diese Art von Studien.

Eine gebräuchliche statistische Methode für diesen Datentyp wird **Sterbetafel** genannt. Die zugrunde liegende Idee der Sterbetafel ist die Unterteilung des Beobachtungszeitraumes in kleinere Zeitintervalle. Es werden alle Personen mit einbezogen, die mindestens über die Länge des Intervalls beobachtet wurden, um die Wahrscheinlichkeit eines terminalen Ereignisses in diesem Intervall zu berechnen. Die geschätzten Wahrscheinlichkeiten für jedes Intervall werden dann verwendet, um die Gesamtwahrscheinlichkeit eines auftretenden Ereignisses zu unterschiedlichen Zeitpunkten zu schätzen.

**Beispiel.** Ist eine neue Nikotinpflastertherapie besser geeignet als eine herkömmliche Nikotinpflastertherapie, um Rauchern das Rauchen abzugewöhnen? Sie könnten eine Studie mit zwei Gruppen von Rauchern durchführen, wobei eine Gruppe mit der herkömmlichen Therapie und die andere mit der experimentellen Therapie behandelt wird. Mithilfe von Sterbetafeln aus den Daten können Sie die gesamte Abstinenzrate zwischen den zwei Gruppen vergleichen und bestimmen, ob die experimentelle Behandlung eine Verbesserung der herkömmlichen Therapie darstellt. Sie können die Überlebensverteilung oder Hazardfunktionen auch grafisch darstellen und diese visuell mit detaillierteren Informationen vergleichen.

**Statistik** Aufgenommene Anzahl, abgehende Anzahl, risikobehaftete Anzahl, Anzahl der terminalen Ereignisse, Anteil der Beendigungen, Anteil der Überlebenden, kumulativer Anteil der Überlebenden (und Standardfehler), Wahrscheinlichkeitsdichte (und Standardfehler), Hazardrate (und Standardfehler) für jedes Zeitintervall jeder Gruppe. Median der Überlebenszeit für jede Gruppe. Wilcoxon-Test (Gehan-Test) zum Vergleichen der Überlebensverteilungen von verschiedenen Gruppen. Diagramme: Funktionsdiagramme für Überleben, Log-Überleben, Dichte, Hazardrate und Eins-minus-Überleben.

Erläuterungen der Daten für Sterbetafeln

**Daten.** Die Zeitvariable sollte quantitativ sein. Die Statusvariable sollte dichotom oder kategorial sowie als ganze Zahl codiert sein und die Ereignisse sollten als einzelne Werte bzw. Bereich von fortlaufenden Werten codiert sein. Faktorvariablen sollten kategorial und als ganze Zahl codiert sein.

**Annahmen.** Wahrscheinlichkeiten für das untersuchte Ereignis dürfen nur von der Zeit nach dem Anfangsereignis abhängen. Sie werden in Bezug auf die absolute Zeit als stabil angenommen. Dies bedeutet, dass die in die Studie zu unterschiedlichen Zeiten aufgenommenen Fälle (wie beispielsweise Patienten mit einer zu unterschiedlichen Zeiten beginnenden Behandlung) sich ähnlich verhalten dürften. Ebenso sollten keine systematischen Differenzen zwischen zensierten und nicht zensierten Fällen auftreten. Wenn zum Beispiel viele der zensierten Fälle Patienten in schlimmeren Zuständen sind, kann das Ihre Ergebnisse beeinflussen.

**Verwandte Prozeduren.** Die Prozedur "Sterbetafeln" verwendet einen versicherungsmathematischen Ansatz für diese Art von Analyse (allgemein als Überlebensanalyse bekannt). Die Prozedur "Kaplan-Meier-Überlebensanalyse" verwendet eine leicht abweichende Methode der Berechnung von Sterbetafeln, die nicht auf einer Aufteilung der Beobachtungsperiode in kleinere Zeitintervalle basiert. Diese Methode empfiehlt sich, wenn Sie nur über eine kleine Anzahl von Beobachtungen verfügen, sodass in jedem Zeitintervall der Überlebensverteilung lediglich wenige Beobachtungen vorliegen. Wenn Sie von Ihren Variablen annehmen, dass sie mit der Überlebenszeit oder mit zu kontrollierenden Variablen (Kovariaten) verbunden sind, verwenden Sie die Prozedur "Cox-Regression". Wenn Kovariaten für den gleichen Fall zu verschiedenen Zeitpunkten unterschiedliche Werte annehmen können, verwenden Sie "Cox-Regression mit zeitabhängigen Kovariaten".

Erstellen von Sterbetafeln

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Überleben > Sterbetafeln...**

2. Wählen Sie eine *numerische* Überlebensvariable aus.

3. Legen Sie die zu untersuchenden Zeitintervalle fest.

4. Wählen Sie eine Statusvariable aus, um die Fälle zu definieren, für die ein terminales Ereignis stattgefunden hat.

5. Klicken Sie auf **Ereignis definieren**, um den Wert der Statusvariablen festzulegen, der ein aufgetretenes Ereignis anzeigt.

Wahlweise können Sie eine Faktorvariable der ersten Ordnung auswählen. Für die Variable der Überlebensverteilung werden versicherungsmathematische Tabellen für jede Kategorie der Faktorvariablen erzeugt.

Sie können auch eine *Faktorvariable* der zweiten Ordnung auswählen. Für die Variable der Überlebensverteilung werden versicherungsmathematische Tabellen für jede Kombination der Faktorvariablen der ersten und zweiten Ordnung erzeugt.

## Sterbetafeln: Bereich für Faktorvariable definieren

Alle Fälle, bei denen die Faktorvariable Werte im festgelegten Bereich annimmt, werden in die Analyse eingeschlossen. Für jeden eindeutigen Wert innerhalb des Bereichs werden getrennte Tabellen (und Diagramme, falls gewünscht) erzeugt.

## Sterbetafeln: Optionen

Sie können verschiedene Aspekte der Analyse von Sterbetafeln beeinflussen.

**Sterbetafel(n).** Um die Anzeige von Sterbetafeln in der Ausgabe zu unterdrücken, inaktivieren Sie das Feld **Sterbetafel(n)**.

**Diagramm.** Hiermit können Sie Diagramme der Überlebensfunktionen anfordern. Wenn Sie Faktorvariable(n) festgelegt haben, werden Diagramme für jede durch die Faktorvariable(n) definierte Untergruppe erzeugt. Verfügbar sind Diagramme für Überleben, Log-Überleben, Hazard, Dichte und "Eins minus Überleben".

- *Überleben.* Zeigt die kumulative Überlebensfunktion auf einer linearen Skala an.
- *Log-Überleben.* Zeigt die kumulative Überlebensfunktion auf einer logarithmischen Skala an.
- *Risiko.* Zeigt die kumulative Hazardfunktion auf einer linearen Skala an.
- *Dichte.* Zeigt die Dichtefunktion an.
- *Eins minus Überleben.* Erzeugt ein Diagramm der Werte "1 - Überlebensfunktion" auf einer linearen Skala.

**Stufen des ersten Faktors vergleichen.** Wenn Sie über eine Kontrollvariable der ersten Ordnung verfügen, können Sie eine der Optionen in diesem Gruppenfeld auswählen, um den Wilcoxon-Test (Gehan-Test) durchzuführen. Dieser Test vergleicht die Überlebensverteilung der Untergruppen. Die Tests werden anhand des Faktors der ersten Ordnung durchgeführt. Falls Sie einen Faktor der zweiten Ordnung festgelegt haben, werden Tests für jede Stufe der Variablen zweiter Ordnung durchgeführt.

## Zusätzliche Funktionen beim Befehl SURVIVAL

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Festlegen mehr als einer abhängigen Variablen.
- Festlegen von unterschiedlichen Zeitintervallen.
- Festlegen mehr als einer Statusvariablen.
- Festlegen von Vergleichen, die nicht alle Faktoren und alle Kontrollvariablen einschließen.
- Berechnen von ungefähren anstatt von genauen Vergleichen.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Parametrische beschleunigte Fehlerzeitmodelle

Eine AFT-Modellanalyse (AFT = Parametric Accelerated Failure Time) ruft die Prozedur für parametrische Überlebensmodelle mit nicht rekursiven Lebenszeitdaten auf. Parametrische Überlebensmodelle gehen

davon aus, dass die Überlebenszeit einer bekannten Verteilung folgt, und diese Analyse passt Modelle für beschleunigte Ausfallzeiten mit ihren Modelleffekten proportional zur Überlebenszeit an.

## Abrufen einer Analyse für parametrische Modelle für beschleunigte Fehlerzeiten

1. Wählen Sie in den Menüs Folgendes aus:

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

2. Wählen Sie eine Quellenvariable aus.

### Zeit

#### Überleben

Einzelne numerische Variable, die die Dauer der Überlebenszeit angibt.

#### Start/Ende

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

### Status

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

#### Fehler/Ereignis

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

#### Rechtszensurierung

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

#### Linkszensurierung

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

#### Intervallzensurierung

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

#### Verfahren für nicht zugeordnete Werte

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

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

### Kovariate(n)

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

### Feste Faktoren

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

### Linkstrunkierung

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

## Parametrische beschleunigte Fehlerzeitmodelle: Kriterien

### Kriterien

Eine optionale Anzeige zur Angabe der allgemeinen Kriterien.

### **Konfidenzintervall**

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

### **Fehlende Werte**

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

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

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

#### **Benutzerdefiniert fehlende Werte sind gültig**

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

### **Statusbehandlung**

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

#### **Datensatz mit Konflikten verwerfen**

Löscht die Datensätze mit Konflikten. Dies ist die Standardeinstellung.

#### **Abrufen der Zeitinformationen nach Status**

Ruft die Zeitinformationen entsprechend dem Status ab

#### **Status anhand der Zeitinformationen ableiten**

Ändert den Status entsprechend den Zeitinformationen.

## **Modelle für parametrische beschleunigte Ausfallzeit: Modell**

### **Modell**

Eine optionale Anzeige zur Angabe der Modelloptionen und -einstellungen.

### **Verteilung der Überlebenszeit**

Eine Option zur Angabe der Verteilung der Überlebenszeit.

#### **Weibull**

Gibt die Weibull-Verteilung an Dies ist die Standardeinstellung.

#### **Exponentiell**

Gibt die Exponentialverteilung an

#### **Logarithmische Normalverteilung**

Gibt die Log-Normal-Verteilung an

#### **Log-logistisch**

Gibt die loglogistische Verteilung an.

### **Kovariateneinstellungen**

Kovariatenvariablen angeben.

### **Faktoreinstellungen**

Geben Sie Faktorvariablen an.

### **Anfangswert des konstanten Terms**

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

### **Anfangswert des Skalenparameters**

Eine Option zur Steuerung der Einstellung des Skalenparameters.

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

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

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

Verwendet den reziproken Wert des Standardfehlers.

### Vom Benutzer angegebener Wert

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

## Parametrische beschleunigte Fehlerzeitmodelle: Schätzung

### Schätzung

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

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

##### Schnell

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

##### Traditionell

Wendet den traditionellen ADMM-Algorithmus an.

#### L-1-Regularisierung anwenden

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

#### Modellkonvergenzkriterien

##### Parameterkonvergenz

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

##### Konvergenz der Zielfunktion

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

##### Konvergenz der Hesse-Matrix

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

#### Konvergenzkriterien für Residuen

Eine Option zur Steuerung des Optimierungsprozesses.

##### Sowohl primales als auch duales Residuum

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

##### Nur primales Residuum

Wendet das Konvergenzkriterium der ursprünglichen Residuen an.

##### Nur duales Residuum

Wendet das doppelte Konvergenzkriterium für Residuen an

#### Methode

Ein optionaler Parameter zur Angabe der Schätzmethode.

### **Automatisch**

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

### **Newton-Raphson**

Wendet die Newton-Raphson-Methode an.

### **L-BFGS**

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

## **Iteration**

### **Maximalzahl der Iterationen**

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

### **Maximale Schritthalbierung**

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

### **Maximale Anzahl von Zeilensuchen**

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

### **Absolute Konvergenz für Iterationsprozess**

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

### **Relative Konvergenz für Iterationsprozess**

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

## **Parametrische beschleunigte Fehlerzeitmodelle: Drucken**

### **Drucken**

Eine optionale Anzeige zur Steuerung der Tabellenausgaben.

### **Details zur Faktorcodierung**

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

### **Den Regressionsparametern zugeordnete Anfangswerte**

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

### **Verlauf der Modelliteration**

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

### **Auswahlergebnisse mit**

Steuert die Anzeige der Details der Featureauswahl.

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

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

### **Nur ausgewählte Variablen**

Nur die ausgewählten Variablen anzeigen.

### Nur nicht ausgewählte Variablen

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

## Parametrische Modelle für beschleunigte Ausfallzeiten: Vorhersagen

### Vorhersage

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

### Zeitwerte für Scoring

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

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

#### Regelmäßige Intervalle

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

#### Zeitdifferenz

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

### Vorhersagen

#### Überleben

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

#### Gefahr

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

#### Kumulierte Gefahr

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

#### Bedingtes Überleben

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

## Modelle für parametrische beschleunigte Ausfallzeit: Diagramm

### Diagramm

#### Diagramm zu Cox-Snell-Residuen

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

#### Funktionsdiagramme

Eine Option zur Steuerung der Funktionsdiagramme.

#### Typ

##### Überleben

Erstellt das Diagramm für Überlebensfunktionen.

**Gefahr**

Erstellt das Diagramm für die Gefahrenfunktionen.

**Dichte**

Erstellt ein Diagramm für die Dichtefunktionen.

**Anzahl anzuzeigender Punkte**

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

**Kovariatenwerte für Diagramm**

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

**Faktorwerte für Diagramm**

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

**Separate Zeilen für**

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

**Maximale Anzahl von Linien in einem Diagramm**

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

## Parametrische beschleunigte Fehlerzeitmodelle: Exportieren

### Exportieren

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

## Survival AFT Ereignisse für Statusvariablen definieren

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

## Parametrische beschleunigte Fehlerzeitmodelle: Kategorie auswählen

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

Kategorie auswählen

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

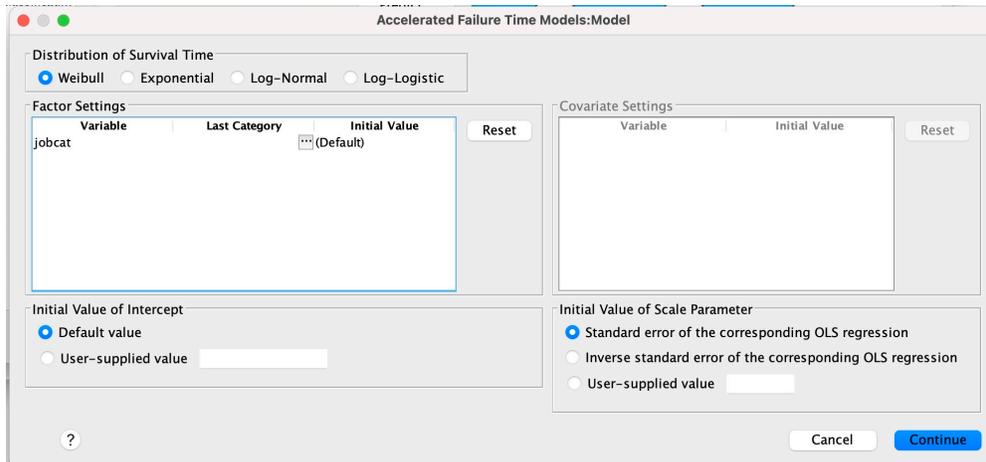


Abbildung 3. Modelle für beschleunigte Lebensdauer-Dialogfenster-Kategorie

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

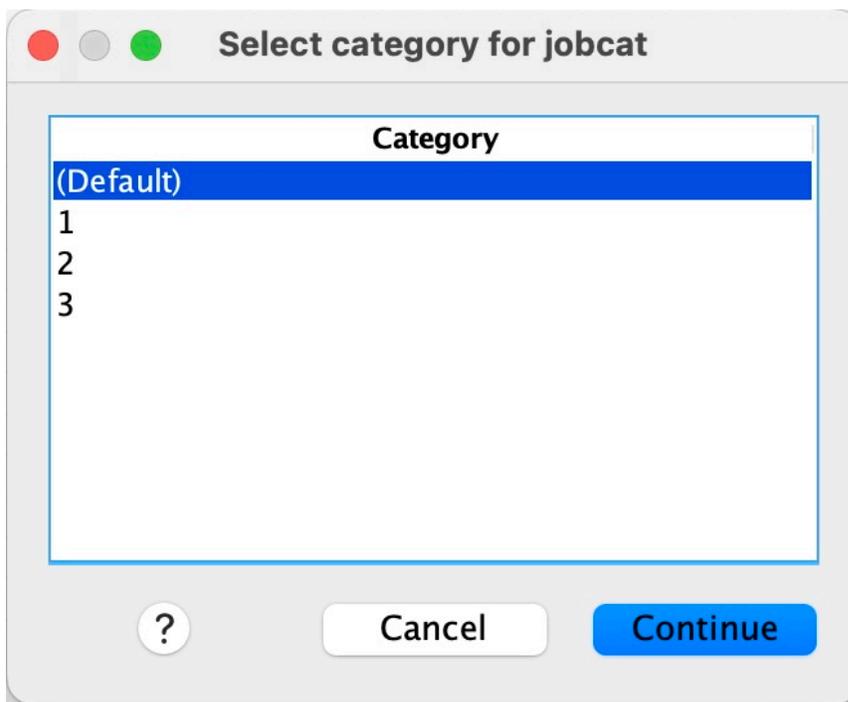


Abbildung 4. Modelle für beschleunigte Lebensdauer-Dialogfenster-Kategorie auswählen

Klicken Sie auf Weiter.

## Parametrische gemeinsam genutzte Frailty-Modelle

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

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

1. Wählen Sie im Menü Folgendes aus:

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

2. Wählen Sie eine Quellenvariable aus.

### Zeit

#### Überleben

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

#### Start/Ende

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

### Thema

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

### Intervallskala

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

### Status

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

#### Fehler/Ereignis

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

#### Rechtszensurierung

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

#### Verfahren für nicht zugeordnete Werte

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

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

### Kovariate(n)

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

### Feste Faktoren

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

## Parametrische gemeinsam genutzte Frailty-Modelle: Kriterien

### Kriterien

Eine optionale Anzeige zur Angabe der allgemeinen Kriterien.

### Konfidenzintervall

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

### **Signifikanzniveau**

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

### **Fehlende Werte**

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

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

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

#### **Benutzerdefiniert fehlende Werte sind gültig**

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

### **Intervallbehandlung**

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

#### **Datensätze mit Konflikten verwerfen**

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

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

Erkennt den Intervallwert aus der Anfangs- und Endzeit.

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

### **Modell**

Eine optionale Anzeige zum Angeben der Modelloptionen und -einstellungen.

### **Verteilung der Überlebenszeit**

Eine Option zur Angabe der Verteilung der Überlebenszeit.

#### **Weibull**

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

#### **Exponentiell**

Gibt die Exponentialverteilung an.

#### **Logarithmische Normalverteilung**

Gibt die Log-Normal-Verteilung an.

#### **Log-logistisch**

Gibt die Log-logistische Verteilung an.

### **Kovariatenereinstellungen**

Kovariatenvariablen angeben.

### **Faktoreinstellungen**

Geben Sie Faktorvariablen an.

### **Anfangswert des konstanten Terms**

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

### **Anfangswert des Skalenparameters**

Eine Option zur Steuerung der Einstellung des Skalenparameters.

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

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

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

Verwendet den reziproken Wert des Standardfehlers.

### **Vom Benutzer angegebener Wert**

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

### **Frailty-Komponente**

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

#### **Gamma**

Gibt die Gammaverteilung an. Dies ist die Standardeinstellung.

#### **Inverse Normalverteilung**

Gibt die inverse Normalverteilung an.

#### **Anfangswert der Varianz**

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

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

### **Schätzung**

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

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

##### **Schnell**

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

##### **Traditionell**

Wendet den traditionellen ADMM-Algorithmus an.

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

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

#### **Modellkonvergenzkriterien**

##### **Parameterkonvergenz**

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

##### **Konvergenz der Zielfunktion**

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

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

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

### **Konvergenzkriterien für Residuen**

Eine Option zur Steuerung des Optimierungsprozesses.

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

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

#### **Nur primales Residuum**

Wendet das Konvergenzkriterium der ursprünglichen Residuen an.

#### **Nur duales Residuum**

Wendet das doppelte Konvergenzkriterium für Residuen an

### **Methode**

Ein optionaler Parameter zur Angabe der Schätzmethode.

#### **Automatisch**

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

#### **Newton-Raphson**

Wendet die Newton-Raphson-Methode an.

#### **L-BFGS**

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

### **Iteration**

#### **Maximalzahl der Iterationen**

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

#### **Maximale Schritthalbierung**

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

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

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

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

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

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

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

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

### **Drucken**

Eine optionale Anzeige, die die Tabellenausgaben steuert.

#### **Details zur Faktorcodierung**

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

#### **Anfangswerte, die den Regressionsparametern zugeordnet sind**

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

## Verlauf der Modelliteration

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

# Parametrische gemeinsam genutzte Frailty-Modelle: Vorhersagen

## Vorhersage

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

### Zeitwerte für Scoring

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

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

#### Regelmäßige Intervalle

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

#### Zeitdifferenz

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

## Vorhersagen

### Überleben

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

### Gefahr

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

### Kumulierte Gefahr

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

### Bedingtes Überleben

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

### Unbedingtes Überleben

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

### Unbedingtes Risiko

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

### Unbedingte cum-Gefahr

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

## Parametrische gemeinsam genutzte Frailty-Modelle: Diagramm

### Diagramm

#### Funktionsdiagramme

Eine Option zur Steuerung der Funktionsdiagramme.

#### Typ

##### Überleben

Erstellt das Diagramm für die unbedingten Überlebensfunktionen.

##### Gefahr

Erstellt das Diagramm für die unbedingten Hazardfunktionen.

##### Dichte

Erstellt ein Diagramm für die Dichtefunktionen.

#### Anzahl anzuzeigender Punkte

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

#### Kovariatenwerte für Diagramm

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

#### Faktorwerte für Diagramm

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

#### Separate Zeilen für

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

#### Maximale Anzahl von Linien in einem Diagramm

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

## Parametrische gemeinsam genutzte Frailty-Modelle: Exportieren

### Exportieren

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

## Parametrische gemeinsam genutzte Frailty-Modelle: Ereignisse definieren

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

1. Wählen Sie im Menü die Option

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

2.

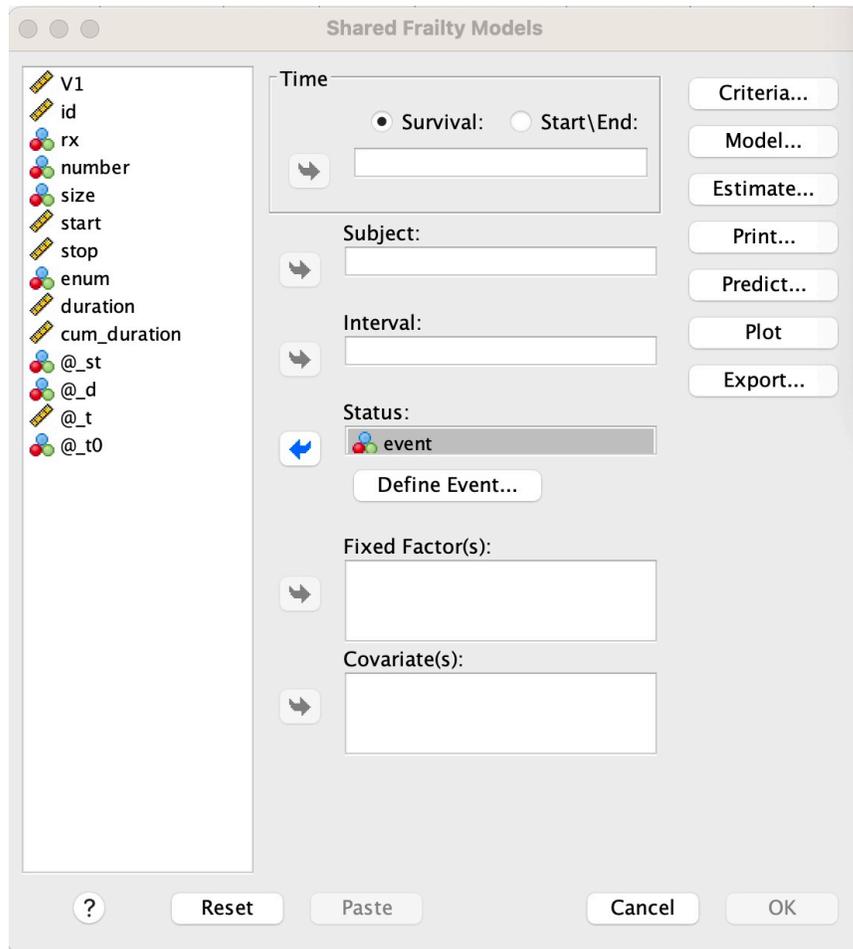


Abbildung 5. Gemeinsam genutzte Frailty-Modelle-Dialogfenster-Status

3.

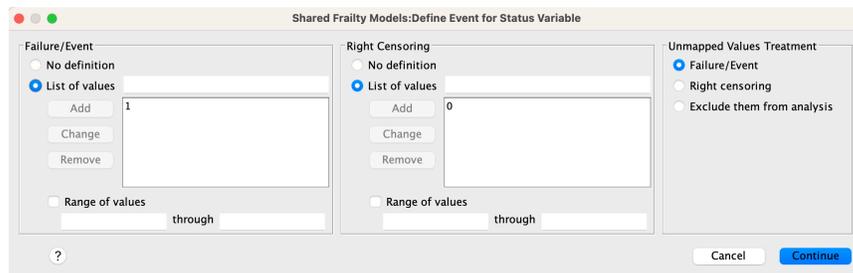


Abbildung 6. Gemeinsam genutzte Fragmentierungsmodelle-Dialogfenster-Status-Ereignis definieren

## Parametrische gemeinsam genutzte Frailty-Modelle-Beispiele

### Beispiel 1

SURVREG RECURRENT Y MIT x1 BY x2

/MODEL SUBJECT = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL.

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

Die Überlebenszeit wird durch eine einzelne Variable y dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

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

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

Alle gültigen Datensätze werden in der Überlebensanalyse verwendet.

### **Beispiel 2**

```
SURVREG RECURRENT Y MIT x1 BY x2
```

```
/MODEL SUBJECT = id FRAILITY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

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

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

Die Subjekte werden durch die Variablen-ID identifiziert.

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

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

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

### **Beispiel 3**

```
SURVREG RECURRENT y1 y2 WITH x1 BY x2(1)
```

```
/MODEL SUBJECT = id FRAILITY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC
```

```
/STATUS VARIABLE=Ereignis FAILURE=1 RIGHT=0.
```

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

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

Die Subjekte werden durch die Variablen-ID identifiziert.

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

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

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

### **Beispiel 4**

```
SURVREG RECURRENT Y MIT x1 BY x2
```

```
/MODEL SUBJECT = ID
```

```
/STATUS VARIABLE=Ereignis FAILURE=1 RIGHT=0
```

```
/PREDICT UNCONDSURVIVAL UNCONDHAZARD UNCONDCUMHAZARD
```

```
/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).
```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von y an eine Kovariate x1 und Faktor x2. Die Überlebenszeit wird durch eine einzelne Variable y dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

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

Die Kurven für unbedingtes oder auf der Bevölkerung basierendes Überleben und Hazard werden durch die Kategorien in x2 getrennt dargestellt.

### Beispiel 5

```
SURVREG RECURRENT Y MIT x1 BY x2
```

```
/MODEL SUBJECT = id FRAILITY=GAMMA DISTRIBUTION=WEIBULL
```

```
/STATUS VARIABLE=Ereignis FAILURE=1 RIGHT=0
```

```
/SCHÄTZUNG HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0 SELECT FEATURES=TRUE  
PENALTY=0.01.
```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von y an eine Kovariate x1 und Faktor x2. Die Überlebenszeit wird durch eine einzelne Variable y dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

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

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

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

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

### Beispiel 6

```
SURVREG RECURRENT Y MIT x1 BY x2
```

```
/MODEL SUBJECT = ID
```

```
/STATUS VARIABLE=infect FAILURE=1 RIGHT=0
```

```
/SCHÄTZUNG MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.
```

Ein parametrisches Überlebensmodell für die gemeinsame Frailty-Anpassung von y an eine Kovariate x1 und Faktor x2. Die Überlebenszeit wird durch eine einzelne Variable y dargestellt.

Die Subjekte werden durch die Variablen-ID identifiziert.

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

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

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

Anwendungsfallname-Nebeneffekt der Behandlung.

Akteure-Ermittler und Praktiker der öffentlichen Gesundheit.

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

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

Patrick möchte eine solche Hypothese bewerten. Die in der Datenstichprobe enthaltenen Variablen werden wie folgt aufgelistet:

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

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

Syntax-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILITY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

**Übersicht:**

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

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

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

## Überlebensanalyse nach Kaplan-Meier

---

Es kann viele Situationen geben, in denen Sie die Verteilung der Zeiten zwischen zwei Ereignissen untersuchen möchten, wie z. B. die Dauer einer Beschäftigung (Zeit zwischen Einstellung und Kündigung eines Arbeitsplatzes). Im Allgemeinen enthalten diese Art von Daten einige zensierte Fälle. Zensierte Fälle sind solche Fälle, für die das zweite Ereignis nicht aufgezeichnet wird (beispielsweise sind Personen nach der Studie weiterhin für das Unternehmen tätig). Die Prozedur "Kaplan-Meier" stellt eine Methode dar, um Modelle, die die Zeit bis zum Eintreten des Ereignisses angeben, in Gegenwart zensierter Fälle zu schätzen. Das Kaplan-Meier-Modell basiert auf der Schätzung bedingter Wahrscheinlichkeiten zu jedem Zeitpunkt eines auftretenden Ereignisses und auf der Bildung des Produktgrenzwerts dieser Wahrscheinlichkeiten zur Schätzung der Überlebensrate zu jedem Zeitpunkt.

**Beispiel.** Ergibt sich aus einer neuen Behandlung gegen AIDS ein therapeutischer Nutzen hinsichtlich der Lebenserwartung? Sie könnten mit zwei Gruppen von AIDS-Patienten eine Studie durchführen, wobei die eine Gruppe eine traditionelle Therapie und die andere die zu erprobende Behandlung erhält. Mit der Konstruktion eines Kaplan-Meier-Modells zu den Daten könnten Sie die Gesamtüberlebensraten zwischen den beiden Gruppen vergleichen, um zu bestimmen, ob die zu erprobende Behandlung eine Verbesserung gegenüber der traditionellen Therapie darstellt. Sie können die Überlebensverteilung oder Hazardfunktionen auch grafisch darstellen und diese visuell mit detaillierteren Informationen vergleichen.

**Statistik** Überlebensstabelle, einschließlich Zeit, Status, kumulative Überlebenswahrscheinlichkeit und Standardfehler, kumulative Ereignisse sowie die verbleibende Anzahl; Mittelwert und Median der Überlebenszeit mit Standardfehler und 95-%-Konfidenzintervall. Diagramme: Überlebensverteilung, Hazard, Log-Überlebensdiagramm, Eins-minus-Überleben.

Erläuterungen der Daten für Kaplan-Meier

**Daten.** Die Zeitvariable muss stetig sein, die Statusvariable kann kategorial oder stetig sein und die Faktorvariablen und geschichteten Variablen müssen kategorial sein.

**Annahmen.** Wahrscheinlichkeiten für das untersuchte Ereignis dürfen nur von der Zeit nach dem Anfangsereignis abhängen. Sie werden in Bezug auf die absolute Zeit als stabil angenommen. Dies bedeutet,

dass die in die Studie zu unterschiedlichen Zeiten aufgenommenen Fälle (wie beispielsweise Patienten mit einer zu unterschiedlichen Zeiten beginnenden Behandlung) sich ähnlich verhalten dürften. Ebenso sollten keine systematischen Differenzen zwischen zensierten und nicht zensierten Fällen auftreten. Wenn zum Beispiel viele der zensierten Fälle Patienten in schlimmeren Zuständen sind, kann das Ihre Ergebnisse beeinflussen.

**Verwandte Prozeduren.** Die Prozedur "Kaplan-Meier" verwendet eine Methode für die Berechnung von Sterbetafeln, welche die Überlebens- oder Hazardfunktion zum Zeitpunkt jedes Ereignisses schätzt. Die Prozedur "Sterbetafeln" verwendet für die Überlebensanalyse einen versicherungsmathematischen Ansatz, der auf einer Zerlegung des Beobachtungszeitraums in kleinere Zeitintervalle basiert. Dies kann bei umfangreichen Stichproben hilfreich sein. Wenn Sie von Ihren Variablen annehmen, dass sie mit der Überlebenszeit oder mit zu kontrollierenden Variablen (Kovariaten) verbunden sind, verwenden Sie die Prozedur "Cox-Regression". Wenn Kovariaten für den gleichen Fall zu verschiedenen Zeitpunkten unterschiedliche Werte annehmen können, verwenden Sie "Cox-Regression mit zeitabhängigen Kovariaten".

Berechnen einer Überlebensanalyse nach Kaplan-Meier

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Überleben > Kaplan-Meier...**

2. Wählen Sie eine Zeitvariable aus.

3. Wählen Sie eine Statusvariable zum Bestimmen von Fällen aus, für welche das terminale Ereignis aufgetreten ist. Diese Variable kann numerisch oder eine *kurze Zeichenfolge* sein. Klicken Sie anschließend auf **Ereignis definieren**.

Wahlweise können Sie eine Faktorvariable auswählen, um Gruppendifferenzen zu untersuchen. Sie können auch eine geschichtete Variable auswählen, mit der sich für jedes Niveau (Schicht) der Variablen getrennte Analysen erzeugen lassen.

## Kaplan-Meier: Ereignis für Statusvariable definieren

Geben Sie den Wert oder die Werte ein, die anzeigen, dass das terminale Ereignis aufgetreten ist. Sie können einen einzelnen Wert, einen Wertebereich oder eine Liste von Werten eingeben. Die Option "Wertebereich" ist nur verfügbar, wenn die Statusvariable eine numerische Variable ist.

## Kaplan-Meier: Faktorstufen vergleichen

Sie können Statistiken anfordern, um die Gleichheit der Überlebensverteilungen für die verschiedenen Stufen des Faktors zu testen. Es stehen die folgenden Statistiken zur Verfügung: Log-Rang, Breslow-Test und Tarone-Ware-Test. Wählen Sie eine der Möglichkeiten aus, um anzugeben, welche Vergleiche durchgeführt werden sollen: "Zusammengefasst über Schichten", "Für jede Schicht", "Paarweise über Schichten" oder "Paarweise für jede Schicht".

**Linearer Trend für Faktorstufen.** Hiermit können Sie die Faktorstufen auf einen linearen Trend prüfen. Diese Option steht nur für Gesamtvergleiche (statt für paarweise Vergleiche) der Faktorstufen zur Verfügung.

- *Protokollrang.* Ein Test zum Vergleich der Gleichheit der Überlebensverteilungen. Alle Zeitpunkte werden in diesem Test gleich gewichtet.
- *Breslow.* Ein Test zum Vergleich der Gleichheit der Überlebensverteilungen. Die Zeitpunkte werden mit der Anzahl der zu den einzelnen Zeitpunkten gefährdeten Fälle gewichtet.
- *Tarone-Ware.* Ein Test zum Vergleich der Gleichheit der Überlebensverteilungen. Die Zeitpunkte werden mit der Quadratwurzel der Anzahl der zu den einzelnen Zeitpunkten gefährdeten Fälle gewichtet.
- *Zusammengefasst über Schichten.* Vergleicht alle Faktorstufen in einem Einzeltest, um die Gleichheit der Überlebenskurven zu prüfen.
- *Paarweise über Schichten.* Vergleicht jedes unterschiedliche Paar von Faktorstufen. Paarweise Trendtests sind nicht verfügbar.
- *Für jede Schicht.* Führt einen separaten Test auf Gleichheit aller Faktorstufen für jede Schicht durch. Wenn keine Schichtungsvariable vorhanden ist, werden die Tests nicht durchgeführt.

- *Paarweise für jede Schicht.* Vergleicht jedes unterschiedliche Paar von Faktorstufen für jede Schicht. Paarweise Trendtests sind nicht verfügbar. Wenn keine Schichtungsvariable vorhanden ist, werden die Tests nicht durchgeführt.

## Kaplan-Meier: Neue Variablen speichern

Sie können die Informationen aus der Kaplan-Meier-Tabelle als neue Variablen speichern, die dann in nachfolgenden Analysen zum Testen von Hypothesen oder Überprüfen von Annahmen verwendet werden können. Es können die Überlebensverteilung, der Standardfehler der Überlebensverteilung sowie Hazard- und kumulative Ereignisse als neue Variablen gespeichert werden.

- *Überleben.* Schätzung der kumulativen Überlebenswahrscheinlichkeit. In der Standardeinstellung besteht der Variablenname aus dem Präfix "sur\_" und einer laufenden Nummer. Wenn beispielsweise der Name "sur\_1" bereits vorhanden ist, weist Kaplan-Meier der Variablen den Namen "sur\_2" zu.
- *Standardfehler der Überlebensverteilung.* Standardfehler der Schätzung der kumulativen Überlebenswahrscheinlichkeit. In der Standardeinstellung besteht der Variablenname aus dem Präfix "se\_" und einer laufenden Nummer. Wenn beispielsweise der Name "se\_1" bereits vorhanden ist, weist Kaplan-Meier der Variablen den Namen "se\_2" zu.
- *Hazard.* Schätzung der kumulativen Hazardfunktion. In der Standardeinstellung besteht der Variablenname aus dem Präfix "haz\_" und einer laufenden Nummer. Wenn beispielsweise der Name "haz\_1" bereits vorhanden ist, weist Kaplan-Meier der Variablen den Namen "haz\_2" zu.
- *Kumulative Ereignisse.* Kumulative Ereignishäufigkeit, wenn Fälle nach ihren Überlebenszeiten und Statuscodes sortiert sind. In der Standardeinstellung besteht der Variablenname aus dem Präfix "cum\_" und einer laufenden Nummer. Wenn beispielsweise der Name "cum\_1" bereits vorhanden ist, weist Kaplan-Meier der Variablen den Namen "cum\_2" zu.

## Kaplan-Meier: Optionen

Sie können verschiedene Arten von Ausgabe für die Kaplan-Meier-Analyse anfordern.

**Statistik** Für die berechneten Überlebensfunktionen können Überlebensstabelle(n), Mittelwert und Median der Überlebenszeit sowie Quartile ausgegeben werden. Wenn Sie Faktorvariablen eingeschlossen haben, werden für jede Gruppe getrennte Statistiken erzeugt.

**Diagramme.** Die Diagramme ermöglichen eine visuelle Überprüfung der Überlebens-, Eins-minus-Überlebens-, Hazard- und Log-Überlebensfunktionen. Wenn Sie Faktorvariablen eingeschlossen haben, werden die Funktionen für jede Gruppe aufgetragen.

- *Überleben.* Zeigt die kumulative Überlebensfunktion auf einer linearen Skala an.
- *Eins minus Überleben.* Erzeugt ein Diagramm der Werte "1 - Überlebensfunktion" auf einer linearen Skala.
- *Hazard.* Zeigt die kumulative Hazardfunktion auf einer linearen Skala an.
- *Log-Überleben.* Zeigt die kumulative Überlebensfunktion auf einer logarithmischen Skala an.

## Zusätzliche Funktionen beim Befehl KM

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Aktivieren Sie Häufigkeitstabellen, bei denen nachzutragende fehlende Fälle als eine von zensierten Fällen separate Kategorie berücksichtigt werden.
- Legen Sie ungleiche Abstände für den Test auf lineare Trends fest.
- Berechnen Sie Perzentile für die Überlebenszeitvariable, die nicht Quartilen entsprechen.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Cox-Regressionsanalyse

Die Cox-Regression erstellt ein Vorhersagemodell für Daten, die die Zeit bis zum Eintreten des Ereignisses angeben. Das Modell erzeugt eine Überlebensfunktion, die die Wahrscheinlichkeit vorhersagt, mit der das interessierende Ereignis zu einer gegebenen Zeit  $t$  für vorgegebene Werte der Prädiktorvariablen aufgetreten ist. Die Form der Überlebensfunktion und die Regressionskoeffizienten für die Prädiktoren werden aus beobachteten Fällen geschätzt. Das Modell kann dann auf neue Fälle angewendet werden, für die Messungen der Prädiktorvariablen vorliegen. Beachten Sie, dass Informationen aus zensierten Fällen, also diejenigen, bei denen das interessierende Ereignis während der Beobachtungszeit nicht auftritt, der Schätzung des Modells nützlich beiträgt.

**Beispiel.** Bestehen für Männer und Frauen unterschiedliche Risiken, durch Zigarettenrauchen an Lungenkrebs zu erkranken? Mit der Konstruktion eines Cox-Regressionsmodells, bei dem Zigarettenverbrauch (gerauchte Zigaretten pro Tag) und Geschlecht als Kovariaten aufgenommen werden, können Sie Hypothesen zum Einfluss von Geschlecht und Zigarettenverbrauch auf den Zeitpunkt des Auftretens von Lungenkrebs überprüfen.

**Statistik** Für jedes Modell:  $-2LL$ , die Likelihood-Quotienten-Statistik und das Chi-Quadrat insgesamt. Für die Variablen im Modell: Parameterschätzungen, Standardfehler und Wald-Statistiken. Für Variablen, die nicht im Modell enthalten sind: Scorestatistiken und Residuen-Chi-Quadrat.

Erläuterung der Daten für die Cox-Regression

**Daten.** Die Zeitvariable muss eine quantitative Variable sein, die Statusvariable dagegen kann kategorial oder stetig sein. Unabhängige Variablen (Kovariaten) können stetig oder kategorial sein. Kategoriale müssen Dummy- oder indikatorcodierte Variablen sein (die Prozedur enthält eine Option zum automatischen Umcodieren für kategoriale Variablen). Geschichtete Variablen müssen kategorial und als ganze Zahlen oder als kurze Zeichenfolgen codiert sein.

**Annahmen.** Beobachtungen müssen unabhängige Variablen sein und das Hazardverhältnis muss über die Zeit konstant sein. Das Hazardverhältnis von einem Fall zum nächsten darf sich über die Zeit also nicht ändern. Letztere Annahme ist als die **Proportional-Hazards-Annahme** bekannt.

**Verwandte Prozeduren.** Falls die Proportional-Hazards-Annahme ihre Gültigkeit verliert (siehe oben), können Sie die Prozedur "Cox-Regression mit zeitabhängigen Kovariaten" verwenden. Falls keine Kovariaten vorhanden sind oder falls Sie nur über eine kategoriale Kovariate verfügen, können Sterbetafeln oder die Prozedur "Kaplan-Meier" verwendet werden, um die Überlebens- oder Hazardfunktionen für die Stichprobe(n) zu untersuchen. Falls in der Stichprobe keine zensierten Daten vorhanden sind (d. h., in jedem Fall das terminale Ereignis eintrat), können Sie die Prozedur "Lineare Regression" verwenden, um das Verhältnis zwischen Prädiktoren und Zeitpunkt bis zum Eintreten des Ereignisses zu modellieren.

Berechnen einer Cox-Regressionsanalyse

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Überleben > Cox-Regression...**

2. Wählen Sie eine Zeitvariable aus. Fälle, deren Zeitwerte negativ sind, werden nicht analysiert.

3. Wählen Sie eine Statusvariable aus und klicken Sie anschließend auf **Ereignis definieren**.

4. Wählen Sie mindestens eine Kovariate aus. Wenn Sie Interaktionsbedingungen einschließen möchten, wählen Sie alle in der Interaktion beteiligten Variablen aus und klicken Sie dann auf **>a\*b>**.

Wahlweise können Sie getrennte Modelle für unterschiedliche Gruppen berechnen, indem Sie eine geschichtete Variable definieren.

### Cox-Regression: Kategoriale Kovariaten definieren

Sie können im einzelnen festlegen, wie bei der Prozedur "Cox-Regression" mit kategorialen Variablen verfahren wird:

**Kovariaten.** Listet alle im Hauptdialogfeld bestimmten Kovariaten auf, 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 Prädiktorvariablen mit Ausnahme der ersten Kategorie wird mit der vorangehenden 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.

## Cox-Regression: Diagramme

Diagramme können ein Hilfsmittel zur Bewertung des geschätzten Modells und zur Interpretation der Ergebnisse sein. Die Überlebens-, Hazard- und Log-minus-Log-Funktionen sowie "Eins minus Überleben" können grafisch dargestellt werden.

### Diagramme

#### Überleben

Zeigt die kumulative Überlebensfunktion auf einer linearen Skala an.

#### Gefahr

Zeigt die kumulative Hazardfunktion auf einer linearen Skala an.

#### Log minus Log

Die kumulative Überlebensschätzung nach Anwendung der  $\ln(-\ln)$ -Transformation auf die Schätzung.

#### Eins-minus-Überleben

Erzeugt ein Diagramm der Werte "1 - Überlebensfunktion" auf einer linearen Skala.

### Kovariatenwerte darstellen bei

Wenn Sie eine kategoriale Kovariante in das Textfeld **Separate Linien für** verschieben, wird für jeden Wert dieser Kovariaten eine separate Linie dargestellt. Diese Option ist nur für kategoriale Kovariaten verfügbar, die in der Liste **Kovariatenwerte darstellen bei** durch ein dem Namen nachgestelltes **(Kat)** gekennzeichnet sind.

### Wert ändern

Weil diese Funktionen von den Werten der Kovariaten abhängen, müssen beim Auftragen der Funktionen über Zeit konstante Werte für die Kovariaten verwendet werden. Die Standardeinstellung sieht vor, dass als konstanter Wert der Mittelwert jeder Kovariaten verwendet wird. Im Gruppenfeld **Wert ändern** können Sie aber auch einen anderen Wert für das Diagramm eingeben.

## Cox-Regression: Neue Variablen speichern

Sie können verschiedene Ergebnisse der Analyse als neue Variablen speichern. Diese Variablen lassen sich für eine anschließende Analyse verwenden, um Hypothesen oder Annahmen zu überprüfen.

### Modellvariablen speichern

Hiermit haben Sie die Möglichkeit, die Überlebensfunktion und den zugehörigen Standardfehler, Log-minus-Log-Schätzungen, Hazardfunktion, partielle Residuen, DfBeta für die Regression und den linearen Prädiktor  $X \cdot \text{Beta}$  als neue Variablen zu speichern.

#### Überlebensfunktion

Der Wert der kumulativen Überlebensfunktion für eine bestimmte Zeit  $t$  entspricht der Überlebenswahrscheinlichkeit bis zu dieser Zeit.

#### Standardfehler der Überlebensfunktion

Standardfehler der Schätzung der kumulativen Überlebenswahrscheinlichkeit.

#### Log-minus-Log-Überlebensfunktion

Die kumulative Überlebensschätzung nach Anwendung der  $\ln(-\ln)$ -Transformation auf die Schätzung.

#### Hazardfunktion

Speichert die Schätzung für die kumulative Hazardfunktion (auch Cox-Snell-Residuum genannt).

#### Partielle Residuen

Sie können partielle Residuen gegen die Überlebenszeit darstellen, um die Proportional-Hazards-Annahme zu testen. Für jede Kovariate im endgültigen Modell wird eine Variable gespeichert. Partielle Residuen stehen nur für Modelle mit mindestens einer Kovariate zur Verfügung.

#### DfBeta

Geschätzte Änderung eines Koeffizienten, wenn ein Fall entfernt wird. Für jede Kovariate im endgültigen Modell wird eine Variable gespeichert. Differenzen in Beta stehen nur für Modelle mit mindestens einer Kovariate zur Verfügung.

#### $X \cdot \text{Beta}$

Lineare Prädiktorbewertung. Die Summe aus dem Produkt mittelwertzentrierter Kovariatenwerte und deren jeweiligen Parameterschätzungen für jeden Fall.

**Anmerkung:** Wenn Sie Cox mit einer zeitabhängigen Kovariate ausführen, werden nur die Differenzen in Beta gespeichert.

### Modellinformationen in XML-Datei exportieren

Parameterschätzungen werden im XML-Format in die angegebene Datei exportiert. Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden.

## Cox-Regression: Optionen

Sie können verschiedene Aspekte der Analyse und Ausgabe steuern.

**Modellstatistik.** Für die Modellparameter stehen Statistiken wie Konfidenzintervalle für  $\text{Exp}(B)$  und Korrelation der Schätzungen zur Verfügung. Sie können diese Statistiken entweder bei jedem Schritt oder nur beim letzten Schritt anfordern.

**Wahrscheinlichkeit für schrittweise Methode.** Wenn Sie die schrittweise Methode ausgewählt haben, können Sie die Wahrscheinlichkeit entweder für die Aufnahme ins Modell oder den Ausschluss aus dem Modell angeben. Eine Variable wird aufgenommen, wenn das Signifikanzniveau ihres  $F$ -Werts für die Aufnahme kleiner als der Aufnahmewert ist, und eine Variable wird ausgeschlossen, wenn das Signifi-

kanzniveau größer als der Ausschlusswert ist. Der Aufnahmewert muss kleiner als der Ausschlusswert sein.

**Maximalzahl der Iterationen.** Hier können Sie die maximale Anzahl an Iterationen für das Modell angeben. Dadurch wird festgelegt, wie lang die Prozedur nach einer Lösung suchen soll.

**Grundlinienfunktion anzeigen.** Hiermit können Sie die Basishazardfunktion und die kumulative Überlebensverteilung beim Mittelwert der Kovariaten anzeigen lassen. Diese Anzeige ist nicht verfügbar, wenn Sie zeitabhängige Kovariaten angegeben haben.

## Cox-Regression: Ereignis für Statusvariable definieren

Geben Sie den Wert oder die Werte ein, die anzeigen, dass das terminale Ereignis aufgetreten ist. Sie können einen einzelnen Wert, einen Wertebereich oder eine Liste von Werten eingeben. Die Option "Wertebereich" ist nur verfügbar, wenn die Statusvariable eine numerische Variable ist.

## Zusätzliche Funktionen beim Befehl COXREG

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Aktivieren Sie Häufigkeitstabellen, bei denen nachzutragende fehlende Fälle als eine von zensierten Fällen separate Kategorie berücksichtigt werden.
- Wählen Sie eine Referenzkategorie für die Methoden der Abweichungskontraste, einfachen Kontraste und Indikatorkontraste aus, die sich von der ersten und letzten Kategorie unterscheidet.
- Legen Sie ungleiche Kategorieabstände für die polynomiale Kontrastmethode fest.
- Legen Sie zusätzliche Iterationskriterien fest.
- Sie können festlegen, wie mit fehlenden Werten verfahren wird.
- Legen Sie die Namen für gespeicherte Variablen fest.
- Schreiben Sie die Ausgabe in eine externe IBM SPSS Statistics-Systemdatei.
- Speichern Sie Daten für jede Gruppe einer aufgeteilten Datei während der Verarbeitung in einer externen Hilfsdatei zwischen. Dadurch lassen sich beim Ausführen von Analysen mit großen Datensets Speicherkapazitäten freihalten. Bei zeitabhängigen Kovariaten steht diese Möglichkeit nicht zur Verfügung.

Vollständige Syntaxinformationen finden Sie in der *Befehlssyntaxreferenz*.

## Berechnen von zeitabhängigen Kovariaten

---

In bestimmten Situationen möchten Sie ein Cox-Regressionsmodell berechnen, aber die Proportional-Hazards-Annahme gilt nicht. Dies können Änderungen der Hazardverhältnisse über Zeit sein, wenn also die Werte einer (oder mehrerer) Kovariaten zu verschiedenen Zeitpunkten unterschiedlich sind. In solchen Fällen benötigen Sie ein erweitertes Cox-Regressionsmodell, in dem **zeitabhängige Kovariaten** angegeben werden.

Zur Analyse eines solchen Modells müssen Sie zuerst die zeitabhängige Kovariate definieren. Um dies zu erleichtern, steht eine 'Systemvariable' zur Verfügung, die die Zeit darstellt. Diese Variable wird als  $T_.$  bezeichnet. Sie können diese Variable verwenden, um zeitabhängige Kovariaten auf zwei allgemeine Arten zu definieren:

- Um die Annahme der proportionalen Hazards in Bezug auf eine bestimmte Kovariate zu testen oder ein erweitertes Cox-Regressionsmodell zu schätzen, das nicht proportionale Hazards zulässt, müssen Sie die zeitabhängige Kovariate als Funktion der Zeitvariablen  $T_.$  und der betreffenden Kovariate definieren. Ein geläufiges Beispiel wäre das einfache Produkt aus Zeitvariable und Kovariate, aber es können auch komplexere Funktionen festgelegt werden. Das Testen der Signifikanz des Koeffizienten der zeitabhängigen Kovariate zeigt, ob die Proportional-Hazards-Annahme angemessen ist.
- Einige Variablen können unterschiedliche Werte zu unterschiedlichen Zeiträumen haben, sind jedoch nicht systematisch mit der Zeit verknüpft. In solchen Fällen müssen Sie eine **segmentierte zeitabhängige Kovariate** definieren, die mithilfe **logischer Ausdrücke** ausgeführt werden kann. Logische Ausdrücke nehmen bei wahren Aussagen den Wert 1 und bei falschen Aussagen den Wert 0 an. Mithilfe

einer Verkettung von logischen Ausdrücken können Sie die zeitabhängige Kovariate aus einem Set von Messwerten erstellen. Wenn Sie beispielsweise einen Blutdruck haben, der einmal pro Woche für die vier Wochen Ihrer Studie gemessen wird (angegeben als  $BP1$  bis  $BP4$ ), können Sie die zeitabhängige Kovariate als  $(T_ < 1) * BP1 + (T_ > = 1 \ \& \ T_ < 2) * BP2 + (T_ > = 2 \ \& \ T_ < 3) * BP3 + (T_ > = 3 \ \& \ T_ < 4) * BP4$  definieren. Sie stellen fest, dass genau einer der Begriffe in Klammern für jeden spezifischen Fall gleich 1 ist und der Rest gleich 0 ist. Zusammenfassend bedeutet diese Funktion, dass, wenn die Zeit weniger als eine Woche beträgt,  $BP1$  verwendet wird. Wenn die Zeit mehr als eine Woche, aber weniger als zwei Wochen beträgt, verwenden Sie  $BP2$  usw.

Im Dialogfeld "Zeitabhängige Kovariate berechnen" können Sie die Steuerelemente zum Erstellen von Funktionen verwenden, um den Ausdruck für eine zeitabhängige Kovariate zu erstellen, oder Sie können ihn direkt im Textbereich "Name" eingeben. Beachten Sie, dass Zeichenfolgekonstanten in Anführungszeichen oder Apostrophe gesetzt und numerische Konstanten in amerikanischem Format mit einem Punkt als Dezimaltrennzeichen eingegeben werden müssen. Alle resultierenden zeitabhängigen Kovariaten müssen als Kovariaten in Ihr Cox-Regressionsmodell aufgenommen werden.

## Berechnen einer zeitabhängigen Kovariate

1. Wählen Sie im Menü

**Analysieren > Überleben > Cox w/Zeit-Dep Cov**

2. Geben Sie einen Ausdruck für die zeitabhängige Kovariate ein.

3. Wählen Sie **Modell** aus, um mit der Cox-Regression fortzufahren.

Hinweis-Stellen Sie sicher, dass Sie die neuen Variablen, die als Kovariaten hinzugefügt werden, in Ihr Cox-Regressionsmodell einschließen.

## Codierungsschemata für kategoriale Variablen

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

## Abweichung

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

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

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

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

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

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

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

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

## Einfach

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

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

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

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

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

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

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

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

## Helmert

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

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

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

```
(1/4  1/4  1/4  1/4)
( 1  -1/3  -1/3  -1/3)
( 0   1  -1/2  -1/2)
( 0   0   1  -1)
```

## Differenz

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

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

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

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

```
( 1/4  1/4  1/4  1/4)
(-1   1   0   0)
(-1/2 -1/2  1   0)
(-1/3 -1/3 -1/3  1)
```

## Polynomial

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

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

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

entspricht dem

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

Gleiche Abstände sind jedoch nicht immer erforderlich. Angenommen, *Wirkstf* stellt verschiedene Dosierungen eines Wirkstoffs dar, der drei verschiedenen Gruppen verabreicht wurde. Wenn der zweiten Gruppe eine doppelt so hohe Dosierung wie der ersten Gruppe und der dritten Gruppe eine dreimal so hohe Dosierung wie der ersten Gruppe verabreicht wurde, weisen die Behandlungskategorien gleiche Abstände auf. Hierbei eignet sich die folgende, aus aufeinander folgenden Ganzzahlen bestehende Metrik:

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

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

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

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

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

## Wiederholt

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

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

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

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

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

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

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

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

Bei LOGLINEAR müssen Sie Folgendes angeben:

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

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

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

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

## Indikator

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

## Kovarianzstrukturen

Dieser Abschnitt enthält zusätzliche Informationen über Kovarianzstrukturen.

**Ante-Dependenz: 1. Ordnung.** Kovarianzstruktur mit heterogenen Varianzen und heterogenen Korrelationen zwischen benachbarten Elementen. Die Korrelation zwischen zwei nicht benachbarten Elementen ist das Produkt der Korrelationen zwischen den Elementen, die zwischen den relevanten Elementen liegen.

$$\begin{pmatrix}
 \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3 \\
 \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3 \\
 \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3 \\
 \sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2
 \end{pmatrix}$$

**AR(1).** Autoregressive Struktur erster Ordnung mit homogenen Varianzen. Die Korrelation zwischen den beiden Elementen entspricht Rho für benachbarte Elemente,  $\text{Rho}^2$  für Elemente, die durch ein Drittel getrennt sind, ist so eingeschränkt, dass  $-1 < \text{Rho} < 1$ .

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

**AR(1): Heterogen.** Autoregressive Struktur erster Ordnung mit heterogenen Varianzen. Die Korrelation zwischen zwei Elementen entspricht  $r$  für benachbarte Elemente,  $r^2$  für zwei Elemente, die durch ein Drittel voneinander getrennt sind, usw. ist so eingeschränkt, dass sie zwischen  $-1$  und  $1$  liegt.

$$\begin{pmatrix}
 \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3 \\
 \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2 \\
 \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\
 \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2
 \end{pmatrix}$$

**ARMA(1,1).** Autoregressive Struktur mit gleitendem Durchschnitt erster Ordnung. Sie hat homogene Varianz. Die Korrelation zwischen zwei Elementen ist identisch mit  $\phi$  für benachbarte Elemente,  $\phi^2$  für Elemente, die durch ein Drittel voneinander getrennt sind, usw. und sind die autoregressiven und gleitenden durchschnittlichen Parameter, und ihre Werte liegen zwischen  $-1$  bis einschließlich  $1$ .

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

**Zusammengesetzte Symmetrie.** Struktur mit konstanter Varianz und konstanter Kovarianz.

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

**Zusammengesetzt symmetrisch: Korrelationsmetrik.** Kovarianzstruktur mit homogenen Varianzen und homogenen Korrelationen zwischen den Elementen.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$

**Zusammengesetzt symmetrisch: heterogen.** Kovarianzstruktur mit heterogenen Varianzen und konstanten Korrelationen zwischen den Elementen.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho$
$(\sigma_2\sigma_1\rho$	$\sigma_2^2$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho$
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	$\sigma_3^2$	$\sigma_4\sigma_3\rho$
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	$\sigma_4^2$

**Diagonal.** Kovarianzstruktur mit heterogenen Varianzen und Korrelation zwischen den Elementen gleich Null.

$(\sigma_1^2$	0	0	0)
(0	$\sigma_2^2$	0	0)
(0	0	$\sigma_3^2$	0)
(0	0	0	$\sigma_4^2$ )

**Direktes Produkt, AR1 (UN\_AR1).** Gibt das Kronecker-Produkt aus einer unstrukturierten Matrix und der anderen Kovarianzmatrix mit Autoregression erster Ordnung an. Die erste unstrukturierte Matrix modelliert die multivariate Beobachtung und die zweite Autoregressionskovarianzstruktur der ersten Ordnung modelliert die Datenkovarianz nach der Zeit oder nach einem anderen Faktor.

**Direktes Produkt unstrukturiert (UN\_UN).** Gibt das Kronecker-Produkt aus zwei unstrukturierten Matrizen an, wobei die erste die multivariate Beobachtung und die zweite die Datenkovarianz über die Zeit oder einen anderen Faktor modelliert.

**Direktes Produkt, zusammengesetzte Symmetrie (UN\_CS).** Gibt das Kronecker-Produkt einer unstrukturierten Matrix und der anderen Kovarianzmatrix mit zusammengesetzter Symmetrie bei konstanter Varianz und Kovarianz an. Die erste unstrukturierte Matrix modelliert die multivariate Beobachtung und die zweite Kovarianzstruktur mit zusammengesetzter Symmetrie modelliert die Datenkovarianz nach der Zeit oder nach einem anderen Faktor.

**Faktoranalytisch: 1. Ordnung** Kovarianzstruktur mit heterogenen Varianzen, die sich aus einem zwischen den Elementen heterogenen Term und einem zwischen den Elementen homogenen Term zusammensetzen. Die Kovarianz zwischen zwei Elementen entspricht der Wurzel aus dem Produkt der heterogenen Varianzanteile.

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

**Faktoranalytisch: 1. Ordnung, heterogen.** Kovarianzstruktur mit heterogenen Varianzen, die sich aus zwei Termen zusammensetzen, die heterogen zwischen den Elementen sind. Die Kovarianz zwischen zwei beliebigen Elementen entspricht der Quadratwurzel aus dem Produkt der ersten ihrer heterogenen Varianzterme.

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

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

**Huynh-Feldt.** "Zirkuläre" Matrix, in der die Kovarianz zwischen zwei beliebigen Elementen gleich dem Durchschnitt ihrer Varianzen minus einer Konstanten ist. Weder die Varianzen noch die Kovarianzen sind konstant.

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

**Skalierte Identität** Struktur mit konstanter Varianz. Es wird angenommen, dass es zwischen den Elementen keine Korrelationen gibt.

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

**Räumlich: Exponent.** Diese Kovarianzstruktur weist homogene Varianzen und heterogene Korrelationen zwischen Elementen auf.  $d_{ij}$  ist der geschätzte Euklidean-Abstand zwischen der Messung  $i^{\text{th}}$  und  $j^{\text{th}}$ .

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

**Räumlich: Exponentiell.** Diese Kovarianzstruktur weist homogene Varianzen und heterogene Korrelationen zwischen Elementen auf.  $d_{ij}$  ist der geschätzte Euklidean-Abstand zwischen der Messung  $i^{\text{th}}$  und  $j^{\text{th}}$ .

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

**Räumlich: gaußsch.** Diese Kovarianzstruktur weist homogene Varianzen und heterogene Korrelationen zwischen Elementen auf.  $d_{ij}$  ist der geschätzte Euklidean-Abstand zwischen der Messung  $i^{\text{th}}$  und  $j^{\text{th}}$ .

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

**Räumlich: linear.** Diese Kovarianzstruktur weist homogene Varianzen und heterogene Korrelationen zwischen Elementen auf.  $d_{ij}$  ist der geschätzte Euklidean-Abstand zwischen der Messung  $i^{\text{th}}$  und  $j^{\text{th}}$  und  $1_{ij}$  ist eine Indikatorfunktion, bei der es sich um 1 handelt, wenn  $pd_{ij} \leq 0$  und 0 anderweitig verwendet wird.

$(\sigma^2$	$\sigma^2(1 - \rho d_{12}) 1_{12}$	$\sigma^2(1 - \rho d_{13}) 1_{13}$	$\sigma^2(1 - \rho d_{14}) 1_{14}$ )
$(\sigma^2(1 - \rho d_{12}) 1_{12}$	$\sigma^2$	$\sigma^2(1 - \rho d_{23}) 1_{23}$	$\sigma^2(1 - \rho d_{24}) 1_{24}$ )

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

**Räumlich: linear logistisch.** Diese Kovarianzstruktur weist homogene Varianzen und heterogene Korrelationen zwischen Elementen auf.  $d_{ij}$  ist der geschätzte Euklidean-Abstand zwischen der Messung  $i^{\text{th}}$  und  $j^{\text{th}}$  und  $1_{ij}$  ist eine Indikatorfunktion, bei der es sich um 1 handelt, wenn  $\rho \log(d_{ij}) \leq 0$  und 0 anderweitig ist.

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

**Räumlich: kugelförmig.** Diese Kovarianzstruktur weist homogene Varianzen und heterogene Korrelationen zwischen Elementen auf.  $r_{ij} = d_{ij}/\rho$ , wobei  $d_{ij}$  der geschätzte Euklidean-Abstand zwischen der Messung  $i^{\text{th}}$  und  $j^{\text{th}}$  ist.  $1_{ij}$  ist eine Indikatorfunktion, bei der es sich um 1 handelt, wenn  $d_{ij} \leq \rho$  und 0 anderweitig ist.

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

**Toeplitz.** Kovarianzstruktur mit homogenen Varianzen und heterogenen Korrelationen zwischen den Elementen. Die Korrelation zwischen benachbarten Elementen ist homogen; die Korrelation zwischen Elementen im Abstand 2 ist wiederum homogen usw.

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

**Toeplitz: Heterogen.** Kovarianzstruktur mit heterogenen Varianzen und heterogenen Korrelationen zwischen Elementen. Die Korrelation zwischen benachbarten Elementen ist homogen; die Korrelation zwischen Elementen im Abstand 2 ist wiederum homogen usw.

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

**Unstrukturiert.** Ganz allgemeine Kovarianzmatrix.

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\
 (\sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\
 (\sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\
 (\sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2)
 \end{array}$$

**Unstrukturiert: Korrelationsmetrik.** Kovarianzstruktur mit heterogenen Varianzen und heterogenen Korrelationen.

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

**Varianzkomponenten.** Struktur, die jedem der angegebenen Zufallseffekte eine skalierte Identitätsstruktur (ID-Struktur) zuweist.

## Bayessche Statistik

IBM SPSS Statistics bietet Unterstützung für die folgenden Bayes-Statistiken.

### t-Tests mit einer und zwei Stichproben

Die Prozedur "Bayes-Inferenz bei einer Stichprobe" bietet Optionen für die Erstellung einer Bayes-Inferenz bei einer Stichprobe und bei T-Test bei zwei Stichproben mit paarigen Werten durch die Charakterisierung von A-posteriori-Verteilungen. Wenn Sie normale Daten haben, können Sie eine normale A-priori-Verteilung verwenden, um eine normale A-posteriori-Verteilung zu erhalten.

### Tests mit binomialem Anteil

Die Prozedur "Bayes-Inferenz bei einer Stichprobe: Binomial" bietet Optionen für die Ausführung der Bayes-Inferenz bei einer Stichprobe bei Binomial-Verteilung. Der interessante Parameter ist  $\pi$ , der die Erfolgswahrscheinlichkeit bei einer festen Anzahl Versuche angibt, die erfolgreich sein oder fehlschlagen können. Beachten Sie, dass die einzelnen Versuche von einander unabhängig sind und die Wahrscheinlichkeit  $\pi$  bei jedem Versuch gleich bleibt. Eine binomiale Zufallsvariable kann als Summe einer festen Anzahl unabhängiger Bernoulli-Versuche aufgefasst werden.

### Analyse mit Poisson-Verteilung

Die Prozedur "Bayes-Inferenz bei einer Stichprobe: Poisson" bietet Optionen für die Ausführung der Bayes-Inferenz bei einer Stichprobe bei Poisson-Verteilung. Bei der Poisson-Verteilung, einem nützlichen Modell für seltene Ereignisse, wird angenommen, dass in kleinen Zeitintervallen die Wahrscheinlichkeit, dass ein Ereignis auftritt, proportional zur Länge der Wartezeit ist. Eine konjugierte A-priori-Verteilung innerhalb der Familie der Gammaverteilungen wird verwendet, wenn die bayessche statistische Inferenz für die Poisson-Verteilung gezeichnet wird.

### Verbundene Stichproben

Das Design der Bayes-Inferenz bei verbundenen Stichproben ähnelt dem der Bayes-Inferenz bei einer Stichprobe hinsichtlich der Handhabung von Stichproben mit paarigen Werten. Sie können die Variablennamen paarweise angeben und die Bayes-Analyse für die Mittelwertdifferenz ausführen.

### t-Tests bei unabhängigen Stichproben

Die Prozedur "Bayes-Inferenz bei unabhängigen Stichproben" bietet Optionen für die Verwendung einer Gruppenvariablen, um zwei nicht zusammengehörige Gruppen zu definieren, und stellt die Bayes-Inferenz für die Differenz der beiden Gruppenmittelwerte bereit. Sie können die Bayes-Faktoren mithilfe verschiedener Methoden schätzen und außerdem die A-posteriori-Verteilung unter der Annahme charakterisieren, dass die Varianzen bekannt oder unbekannt sind.

### Paarweise Korrelation (Pearson)

Die bayessche Inferenz über Pearson-Korrelationskoeffizienten misst die lineare Beziehung zwischen zwei metrischen Variablen, die gemeinsam einer bivariaten Normalverteilung folgen. Die konventionelle statistische Inferenz über Korrelationskoeffizienten wurde umfassend diskutiert und das Verfahren wird seit langem in IBM SPSS Statistics angeboten. Mit dem Design der bayesschen Inferenz über Pearson-Korrelationskoeffizienten können Sie die Bayes-Inferenz zeichnen, indem Sie Bayes-Faktoren schätzen und A-posteriori-Verteilungen charakterisieren.

### Lineare Regression

Die bayessche Inferenz über Pearson-Korrelation ist eine statistische Methode, die im Wesentlichen bei der quantitativen Modellierung verwendet wird. Lineare Regression ist ein grundlegendes Stan-

dardkonzept, bei dem Wissenschaftler Werte eines metrischen Ergebnisses mithilfe der Werte von mehreren Variablen erläutern oder vorhersagen. Bayessche univariate lineare Regression ist ein Konzept für die lineare Regression, bei dem die statistische Analyse im Rahmen der bayesschen Inferenz durchgeführt wird.

### **Einfaktorielle ANOVA**

Die bayessche einfaktorielle ANOVA-Prozedur erstellt eine einfaktorielle Varianzanalyse für eine quantitativ abhängige Variable durch eine einzelne Faktorvariable (unabhängige Variable). Mit der Varianzanalyse wird die Hypothese überprüft, dass mehrere Mittelwerte gleich sind. SPSS Statistics unterstützt Bayes-Faktoren sowie konjugierte und nicht informative A-priori-Verteilungen.

### **Loglineare Regressionsmodelle**

Das Design für das Testen der Unabhängigkeit zweier Faktoren erfordert zwei kategoriale Variablen für die Erstellung einer Kontingenztabelle und führt eine bayessche Inferenz für die Zeilen/Spaltenzuordnung durch. Sie können die Bayes-Faktoren unter Annahme verschiedener Modelle schätzen und die gewünschte A-posteriori-Verteilung durch Simulieren des simultanen Zuverlässigkeitsintervalls für Interaktionsterme charakterisieren.

### **Einfaktorielle ANOVA mit Messwiederholung**

Die bayessche einfaktorielle ANOVA mit Messwiederholung misst einen Faktor desselben Subjekts zu bestimmten Zeitpunkten oder unter bestimmten Bedingungen und Subjekte dürfen sich innerhalb der Messniveaus überkreuzen. Es wird angenommen, dass jedes Subjekt eine einzelne Beobachtung zu jedem Zeitpunkt und jeder Bedingung hat (die Interaktion bei der Behandlung des Subjekts wird nicht berücksichtigt).

## **Bayes-Inferenz bei einer Stichprobe: Normal**

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Die Prozedur Bayes-Inferenz bei einer Stichprobe: Normal stellt Optionen für das Durchführen der bayesschen Inferenz für paarige t-Tests mit einer und zwei Stichproben durch Charakterisierung von A-posteriori-Verteilungen bereit. Wenn Sie normale Daten haben, können Sie eine normale A-priori-Verteilung verwenden, um eine normale A-posteriori-Verteilung zu erhalten.

1. Wählen Sie in den Menüs Folgendes aus:

**Analysieren > Bayes-Statistik > Eine Stichprobe normal**

2. Wählen Sie die entsprechenden **Testvariablen** in der Liste **Verfügbare Variablen** aus. Es muss mindestens eine Variable ausgewählt werden.

**Anmerkung:** Die Liste mit verfügbaren Variablen enthält alle Variablen außer Datums- und Zeichenfolgevariablen.

3. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 2. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

1

2

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.
4. Wählen Sie die entsprechenden Einstellungen für **Werte für Datenvarianz und Hypothese** aus und/oder geben Sie sie ein. Die Tabelle gibt die Variablen wieder, die zurzeit in der Liste **Testvariablen** enthalten sind. Sobald der Liste **Testvariablen** Variablen hinzugefügt oder daraus entfernt werden, werden diese Variablen automatisch den Variablenpalten der Tabelle hinzugefügt oder daraus entfernt.

- Wenn die Liste **Testvariablen** mindestens eine Variable enthält, werden die Spalten **Varianz bekannt** und **Varianzwert** aktiviert.

#### **Varianz bekannt**

Wählen Sie diese Option für jede Variable aus, wenn die Varianz bekannt ist.

#### **Varianzwert**

Ein optionaler Parameter, der den Varianzwert für beobachtete Daten angibt, sofern dieser bekannt ist.

- Wenn die Liste **Testvariablen** mindestens eine Variable enthält und **A-posteriori-Verteilung charakterisieren** nicht ausgewählt ist, werden die Spalten **Nulltestwert** und **g-Wert** aktiviert.

#### **Nulltestwert**

Ein erforderlicher Parameter, der den Nullwert in der Bayes-Faktorschätzung angibt. Nur ein Wert ist zulässig und 0 ist der Standardwert.

#### **g-Wert**

Gibt den Wert für die Definition von  $\psi^2 = g\sigma_x^2$  in der Bayes-Faktorschätzung an. Wenn der **Varianzwert** angegeben wird, nimmt der **Wert** standardmäßig den Wert 1 an. Wenn der **Varianzwert** nicht angegeben ist, können Sie einen festen **g** angeben oder den Wert weglassen, um ihn zu integrieren.

5. Optional können Sie auf **Kriterien** klicken, um Einstellungen für „Bayes-Inferenz bei einer Stichprobe: Kriterien“ auf Seite 114“ (Zuverlässigkeitsintervall in Prozent, Optionen für fehlende Werte und Einstellungen für numerische Methoden) anzugeben, oder klicken Sie auf **A-priori-Wahrscheinlichkeiten**,

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

<sup>2</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

um Einstellungen für „Bayes-Inferenz bei einer Stichprobe: Normale A-priori-Wahrscheinlichkeiten“ auf Seite 115“ (Typ der A-priori-Wahrscheinlichkeiten, beispielsweise Inferenzparameter, und Mittelwert für angegebene Varianz oder Genauigkeit) anzugeben.

## Bayes-Inferenz bei einer Stichprobe: Kriterien

Sie können die folgenden Analyse Kriterien für die Bayes-Inferenz bei einer Stichprobe angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

### Fehlende Werte

Geben Sie die Methode zum Handhaben von fehlenden Werte an.

### Paarweiser Fallausschluss

Dies ist die Standardeinstellung. Sie schließt Datensätze mit fehlenden Werten auf einer Analyse-für-Analyse-Basis aus. Datensätze mit fehlenden Werten für ein Feld, das für einen bestimmten Test verwendet wird, werden aus diesem Test ausgeschlossen.

### Fälle listenweise ausschließen

Diese Einstellung schließt Datensätze mit fehlenden Werten listenweise aus. Datensätze mit fehlenden Werten für ein beliebiges Feld, das in einem beliebigen Unterbefehl genannt wird, werden aus allen Analysen ausgeschlossen.

**Anmerkung:** Die folgenden Optionen sind nur verfügbar, wenn die Option **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** für **Bayes-Analyse** ausgewählt ist.

### Numerische Methode

Geben Sie die numerische Methode an, die zum Schätzen des Integrals verwendet wird.

### Adaptive Gauss-Lobatto-Quadratur

Dies ist die Standardeinstellung. Sie ruft die Methode der adaptiven Gauss-Lobatto-Quadratur auf.

### Toleranz

Geben Sie den Toleranzwert für die numerischen Methoden an. Die Standardeinstellung ist 0,000001. Die Option ist nur verfügbar, wenn die Einstellung **Adaptive Gauss-Lobatto-Quadratur** ausgewählt ist.

### Maximalzahl der Iterationen

Geben Sie die maximale Anzahl der Iterationen der Methode der adaptiven Gauss-Lobatto-Quadratur an. Dieser Wert muss eine positive Ganzzahl sein. Die Standardeinstellung ist 2000. Die Option ist nur verfügbar, wenn die Einstellung **Adaptive Gauss-Lobatto-Quadratur** ausgewählt ist.

### Monte-Carlo-Approximation

Diese Option ruft die Methode der Monte-Carlo-Approximation auf.

### Benutzerdefinierten Startwert festlegen

Bei Auswahl dieser Option können Sie im Feld **Start** einen benutzerdefinierten Startwert angeben.

### Startwert

Geben Sie einen Startwert für Zufallszahlen für die Methode der Monte-Carlo-Approximation an. Dieser Wert muss eine positive Ganzzahl sein. Standardmäßig wird ein Startwert für Zufallszahlen zugewiesen.

### Anzahl der Monte-Carlo-Stichproben

Geben Sie die Anzahl der Punkte an, für die Stichproben für die Monte-Carlo-Approximation gezogen werden sollen. Dieser Wert muss eine positive Ganzzahl sein. Der Standardwert ist 1.000.000. Die Option ist nur verfügbar, wenn die Einstellung **Monte-Carlo-Approximation** ausgewählt ist.

## Bayes-Inferenz bei einer Stichprobe: Normale A-priori-Wahrscheinlichkeiten

Sie können die folgenden A-priori-Verteilungskriterien für die Bayes-Inferenz bei einer Stichprobe angeben:

**Anmerkung:** Viele Anwendungsforscher stellen möglicherweise die Notwendigkeit zum Angeben einer A-priori-Wahrscheinlichkeit in Frage. A-priori-Wahrscheinlichkeiten als Referenz minimieren das Problem, wenn die A-priori-Wahrscheinlichkeit bei zunehmenden Daten überlastet ist. Wenn A-priori-Wahrscheinlichkeiten als Informationen angegeben werden, können Bayes-Methoden diese Informationen effektiv nutzen. Die Anforderung zum Angeben einer A-priori-Wahrscheinlichkeit sollte nicht als Hinderungsgrund für die Verwendung einer Bayes-Analyse betrachtet werden.

### A-priori-Wahrscheinlichkeit für Varianz/Genauigkeit

Stellt Optionen zum Definieren von Varianz- und Genauigkeitswerten bereit.

#### Varianz

Wählen Sie diese Option aus, um die A-priori-Verteilung für den Varianzparameter anzugeben.

Wenn diese Option ausgewählt ist, enthält die Liste **A-priori-Verteilung** die folgenden Optionen:

**Anmerkung:** Wenn die Datenvarianz für einige Variablen bereits definiert ist, werden die folgenden Einstellungen für diese Variablen ignoriert.

- **Diffus** - die Standardeinstellung. Gibt die diffuse A-priori-Verteilung an.
- **Invers Chi-Quadrat** - gibt Verteilung und Parameter für inverse- $\chi^2(v_0, \sigma^2_0)$  an, wobei  $v_0 > 0$  der Freiheitsgrad und  $\sigma^2_0 > 0$  der Skalenparameter ist.
- **Invers Gamma** - gibt Verteilung und Parameter für inverse-Gamma( $\alpha_0, \beta_0$ ) an, wobei  $\alpha_0 > 0$  der Formparameter und  $\beta_0 > 0$  der Skalenparameter ist.
- **Jeffreys S2** - Gibt die nicht aufschlussreiche Version vor  $\propto 1/\sigma^2_0$  an.
- **Jeffreys S4** - gibt die nicht informative A-priori-Verteilung  $1/\sigma^4_0$  an.

#### Genauigkeit

Wählen Sie diese Option aus, um die A-priori-Verteilung für den Genauigkeitsparameter anzugeben. Wenn diese Option ausgewählt ist, enthält die Liste **A-priori-Verteilung** die folgenden Optionen:

- **Gamma** - gibt Verteilung und Parameter für Gamma ( $\alpha_0, \beta_0$ ) an, wobei  $\alpha_0 > 0$  der Formparameter und  $\beta_0 > 0$  der Skalenparameter ist.
- **Chi-Quadrat** - gibt Verteilung und Parameter für  $\chi^2(v_0)$  an, wobei  $v_0 > 0$  der Freiheitsgrad ist.

#### Formparameter

Geben Sie den Formparameter  $a_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist.

#### Skalenparameter

Geben Sie den Skalenparameter  $b_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist. Je größer der Skalenparameter ist, desto mehr verteilt sich die Verteilung.

### A-priori-Verteilung für Mittelwert für angegebene Varianz/Genauigkeit

Geben Sie die A-priori-Verteilung für den bezüglich der Varianz bedingten Mittelwertparameter oder den Genauigkeitsparameter an.

#### Normal

Gibt die Verteilung und die Parameter für Normal( $\mu_0, K^{-1}_0\sigma^2_0$ ) für die Varianz oder Normal( $\mu_0, K_0/\sigma^2_0$ ) für die Genauigkeit an, wobei  $\mu_0 \in (-\infty, \infty)$  und  $\sigma^2 > 0$ .

#### Lageparameter

Geben Sie einen numerischen Wert ein, der den Lageparameter für die Verteilung angibt.

### Skalenparameter

Geben Sie den Skalenparameter  $b_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist.

### Kappa

Geben Sie den Wert von  $K_0$  in  $\text{Normal}(\mu_0, K_0^{-1}\sigma^2_0)$  oder  $\text{Normal}(\mu_0, K_0/\sigma^2_0)$  an. Sie müssen einen einzelnen Wert angeben, der größer als 0 ist (1 ist der Standardwert).

### Diffus

Die Standardeinstellung, die die diffuse A-priori-Verteilung 1 angibt.

## Bayes-Inferenz bei einer Stichprobe: Binomial

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Die Methode "Bayes-Interferenz bei einer Stichprobe: Binomiale Prozedur" bietet Optionen für die Ausführung von Bayes-Interferenz bei Binomial-Verteilung. Der interessante Parameter ist  $\pi$ , der die Erfolgswahrscheinlichkeit bei einer festen Anzahl Versuche angibt, die erfolgreich sein oder fehlschlagen können. Beachten Sie, dass die einzelnen Versuche von einander unabhängig sind und die Wahrscheinlichkeit  $\pi$  bei jedem Versuch gleich bleibt. Eine binomiale Zufallsvariable kann als Summe einer festen Anzahl unabhängiger Bernoulli-Versuche aufgefasst werden.

Es ist zwar nicht erforderlich, aber beim Schätzen eines Binomialparameters wird in der Regel eine A-priori-Verteilung aus der Familie der Betaverteilungen ausgewählt. Die Betafamilie ist ein Konjugat für die Binomialfamilie und führt daher zur A-posteriori-Verteilung mit einem geschlossenen Format, die noch zur Familie der Betaverteilungen gehört.

1. Wählen Sie Folgendes in den Menüs aus:

#### **Analysieren > Bayes-Statistik > Binomial bei einer Stichprobe**

2. Wählen Sie die entsprechenden **Testvariablen** in der Liste **Verfügbare Variablen** aus. Es muss mindestens eine Variable ausgewählt werden.

**Anmerkung:** Die Liste mit verfügbaren Variablen enthält alle Variablen außer Datums- und Zeichenfolgevariablen.

3. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 3. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0

*Tabelle 3. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen (Forts.)*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

3

4

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.
4. Wählen Sie die entsprechenden Einstellungen für **Werte für Erfolgskategorien und Hypothese** aus und/oder geben Sie sie ein. Die Tabelle gibt die Variablen wieder, die zurzeit in der Liste **Testvariablen** enthalten sind. Sobald der Liste **Testvariablen** Variablen hinzugefügt oder daraus entfernt werden, werden diese Variablen automatisch den Variablenpaarspalten der Tabelle hinzugefügt oder daraus entfernt.
- Wenn **A-posteriori-Verteilung charakterisieren** als **Bayes-Analyse** ausgewählt ist, ist die Spalte **Erfolgskategorien** aktiviert.
  - Wenn **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** als **Bayes-Analyse** ausgewählt sind, sind alle bearbeitbaren Spalten aktiviert.

#### **Punkt Null**

Aktiviert und inaktiviert die Option **Nullproportion**. Wenn die Einstellung aktiviert ist, sind die Optionen **Null-A-priori-Formparameter** und **Null-A-priori-Skalenparameter** inaktiviert.

#### **Null-A-priori-Formparameter**

Gibt den Formparameter  $a_0$  unter der Nullhypothese der Binomialverteilung an.

#### **Null-A-priori-Skalenparameter**

Gibt den Skalenparameter  $b_0$  unter der Nullhypothese der Binomialverteilung an.

#### **Nullanteil**

Gibt den Formparameter  $a_0$  und den Skalenparameter  $b_0$  unter der Nullhypothese für eine konjugierte A-priori-Verteilung an (um die Beta- und Haldane-A-priori-Verteilungen zu berücksichtigen). Der gültige Bereich sind numerische Werte zwischen 0 und 1.

#### **Alternativer A-priori-Formparameter**

Ein erforderlicher Parameter für die Angabe von  $a_0$  unter der Alternativhypothese der Binomialverteilung, wenn ein Bayes-Faktor geschätzt werden soll.

#### **Alternativer A-priori-Skalenparameter**

Ein erforderlicher Parameter für die Angabe von  $b_0$  unter der Alternativhypothese der Binomialverteilung, wenn ein Bayes-Faktor geschätzt werden soll.

#### **Erfolgskategorien**

Stellt Optionen zum Definieren von konjugierten A-priori-Verteilungen bereit. Die bereitgestellten Optionen geben an, wie Erfolg für numerische Variablen und Zeichenfolgevariablen definiert ist, wenn die Datenwerte gegen den Testwert getestet werden.

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

<sup>4</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### **Letzte Kategorie**

Dies ist die Standardeinstellung. Der Test auf Binomialverteilung wird unter Verwendung des letzten numerischen Werts ausgeführt, der in der Kategorie gefunden wird, nachdem diese in aufsteigender Folge sortiert wurde.

### **Erste Kategorie**

Der Test auf Binomialverteilung wird unter Verwendung des ersten numerischen Werts ausgeführt, der in der Kategorie gefunden wird, nachdem diese in aufsteigender Folge sortiert wurde.

### **Mittelpunkt**

Verwendet die numerischen Werte  $\geq$  Mittelpunkt als Fälle. Ein Mittelpunktwert ist der Durchschnitt der minimalen und maximalen Stichprobendaten.

### **Trennwert**

Verwendet die numerischen Werte  $\geq$  dem angegebenen Trennwert als Fälle. Die Einstellung muss ein einzelner numerischer Wert sein.

### **Ebene**

Behandelt vom Benutzer angegebene Zeichenfolgewerte (können mehr als 1 sein) als Fälle. Die einzelnen Werte müssen durch Kommas getrennt werden.

5. Optional können Sie auf **Kriterien** klicken, um Einstellungen für „Bayes-Inferenz bei einer Stichprobe: Kriterien“ auf Seite 114“ (Zuverlässigkeitsintervall in Prozent, Optionen für fehlende Werte und Einstellungen für numerische Methoden) anzugeben, oder klicken Sie auf **A-priori-Verteilungen**, um Einstellungen für „Bayes-Inferenz bei einer Stichprobe: Binomiale/Poisson-A-priori-Wahrscheinlichkeiten“ auf Seite 118“ (konjugierte oder benutzerdefinierte A-priori-Verteilungen) anzugeben.

## **Bayes-Inferenz bei einer Stichprobe: Binomiale/Poisson-A-priori-Wahrscheinlichkeiten**

Sie können die folgenden A-priori-Verteilungskriterien für die Bayes-Inferenz bei einer Stichprobe angeben:

**Anmerkung:** Viele Anwendungsforscher stellen möglicherweise die Notwendigkeit zum Angeben einer A-priori-Wahrscheinlichkeit in Frage. A-priori-Wahrscheinlichkeiten als Referenz minimieren das Problem, wenn die A-priori-Wahrscheinlichkeit bei zunehmenden Daten überlastet ist. Wenn A-priori-Wahrscheinlichkeiten als Informationen angegeben werden, können Bayes-Methoden diese Informationen effektiv nutzen. Die Anforderung zum Angeben einer A-priori-Wahrscheinlichkeit sollte nicht als Hinderungsgrund für die Verwendung einer Bayes-Analyse betrachtet werden.

### **Formparameter**

Geben Sie bei binominalen A-priori-Wahrscheinlichkeiten den Formparameter  $a_0$  für die Betaverteilung an.

Geben Sie bei A-priori-Wahrscheinlichkeiten mit Poisson-Verteilung den Formparameter  $a_0$  für Gammaverteilung an.

Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist.

### **Skalenparameter**

Geben Sie bei binomialen A-priori-Wahrscheinlichkeiten den Skalenparameter  $b_0$  für Betaverteilung an.

Geben Sie bei Poisson-A-priori-Wahrscheinlichkeiten den Skalenparameter  $b_0$  für die Gammaverteilung an.

Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist.

## **Bayes-Inferenz bei einer Stichprobe: Poisson**

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Die Methode "Bayesian One Sample Inference: Poisson" bietet Optionen für die Ausführung von Bayesian One-Sample-Inferenz bei Poisson-Verteilung. Bei der Poisson-Verteilung, einem nützlichen Modell für

seltene Ereignisse, wird angenommen, dass in kleinen Zeitintervallen die Wahrscheinlichkeit, dass ein Ereignis auftritt, proportional zur Länge der Wartezeit ist. Eine konjugierte A-priori-Verteilung innerhalb der Familie der Gammaverteilungen wird verwendet, wenn die bayessche statistische Inferenz für die Poisson-Verteilung gezeichnet wird.

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayessche Statistik > Ein Beispiel für Poisson**

2. Wählen Sie die entsprechenden **Testvariablen** in der Liste **Verfügbare Variablen** aus. Es muss mindestens eine Variable ausgewählt werden.

**Anmerkung:** Die Liste mit verfügbaren Variablen enthält alle Variablen außer Datums- und Zeichenfolgevariablen.

3. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 4. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

5

6

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.

4. Wählen Sie die entsprechenden Einstellungen für **Hypothesenwerte** aus und/oder geben Sie sie ein. Die Tabelle gibt die Variablen wieder, die zurzeit in der Liste **Testvariablen** enthalten sind. Sobald der Liste **Testvariablen** Variablen hinzugefügt oder daraus entfernt werden, werden diese Variablen automatisch den Variablenpaarspalten der Tabelle hinzugefügt oder daraus entfernt.

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

<sup>6</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- Wenn **A-posteriori-Verteilung charakterisieren** als **Bayes-Analyse** ausgewählt ist, ist keine der Spalten aktiviert.
- Wenn **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** als **Bayes-Analyse** ausgewählt sind, sind alle bearbeitbaren Spalten aktiviert.

#### **Punkt Null**

Aktiviert und inaktiviert die Option **Nullanteil**. Wenn die Einstellung aktiviert ist, sind die Optionen **Null-A-priori-Formparameter** und **Null-A-priori-Skalenparameter** inaktiviert.

#### **Null-A-priori-Formparameter**

Gibt den Formparameter  $a_0$  unter der Nullhypothese der Poisson-Verteilung an.

#### **Null-A-priori-Skalenparameter**

Gibt den Skalenparameter  $b_0$  unter der Nullhypothese der Poisson-Verteilung an.

#### **Nullrate**

Gibt den Formparameter  $a_0$  und den Skalenparameter  $b_0$  unter der Nullhypothese für eine kombinierte A-priori-Verteilung an (zur Aufnahme der Poisson-Gamma-Beziehung). Der Mindestwert muss ein numerischer Wert sein, der größer als 0 ist. Der Höchstwert muss ein maximaler Doppelwert sein.

#### **Alternativer A-priori-Formparameter**

Ein erforderlicher Parameter für die Angabe von  $a_1$  unter der Alternativhypothese der Poisson-Verteilung, wenn ein Bayes-Faktor geschätzt werden soll.

#### **Alternativer A-priori-Skalenparameter**

Ein erforderlicher Parameter für die Angabe von  $b_1$  unter der Alternativhypothese der Poisson-Verteilung, wenn ein Bayes-Faktor geschätzt werden soll.

5. Optional können Sie auf **Kriterien** klicken, um Einstellungen für "„Bayes-Inferenz bei einer Stichprobe: Kriterien“ auf Seite 114" (Zuverlässigkeitsintervall in Prozent, Optionen für fehlende Werte und Einstellungen für numerische Methoden) anzugeben, oder klicken Sie auf **A-priori-Verteilungen**, um Einstellungen für "„Bayes-Inferenz bei einer Stichprobe: Binomiale/Poisson-A-priori-Wahrscheinlichkeiten“ auf Seite 118" (konjugierte oder benutzerdefinierte A-priori-Verteilungen) anzugeben.

## **Bayes-Inferenz bei verbundenen Stichproben: Normal**

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Das Verfahren "Bayes-Inferenz bei verbundenen Stichprobe: Normal" stellt Optionen für die Bayes-Inferenz bei einer Stichprobe mit paarigen Werten bereit. Sie können die Variablennamen paarweise angeben und die Bayes-Analyse für die Mittelwertdifferenz ausführen.

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayes-Statistik > Normal bei verbundenen Stichproben**

2. Wählen Sie in der Liste **Verfügbare Variablen** die entsprechenden paarigen Variablen aus. Es muss mindestens ein Paar Quellenvariablen ausgewählt werden. Für ein bestimmtes Paarset können nicht mehr als zwei Quellenvariablen ausgewählt werden.

**Anmerkung:** Die Liste mit verfügbaren Variablen enthält alle Variablen außer Zeichenfolgevariablen.

3. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 5. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

7

8

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.
4. Wählen Sie die entsprechenden Einstellungen für **Werte für Datenvarianz und Hypothese** aus und/oder geben Sie sie ein. Die Tabelle gibt die Variablenpaare wieder, die zurzeit in der Liste **Paarige Variablen** enthalten sind. Sobald der Liste **Paarige Variablen** Variablenpaare hinzugefügt oder daraus entfernt werden, werden diese Variablenpaare automatisch den Variablenpaarspalten der Tabelle hinzugefügt oder daraus entfernt.
- Wenn die Liste **Paarige Variablen** mindestens ein Variablenpaar enthält, werden die Spalten **Varianz bekannt** und **Varianzwert** aktiviert.
- Varianz bekannt**  
Wählen Sie diese Option für jede Variable aus, wenn die Varianz bekannt ist.
- Varianzwert**  
Ein optionaler Parameter, der den Varianzwert für beobachtete Daten angibt, sofern dieser bekannt ist.
- Wenn die Liste **Paarige Variablen** mindestens ein Variablenpaar enthält und **A-posteriori-Verteilung charakterisieren** nicht ausgewählt ist, werden die Spalten **Nulltestwert** und **g-Wert** aktiviert.
- Nulltestwert**  
Ein erforderlicher Parameter, der den Nullwert in der Bayes-Faktorschätzung angibt. Nur ein Wert ist zulässig und 0 ist der Standardwert.
- g-Wert**  
Gibt den Wert für die Definition von  $\psi^2 = g\sigma^2_x$  in der Bayes-Faktorschätzung an. Wenn der **Varianzwert** angegeben wird, nimmt der **Wert** standardmäßig den Wert 1 an. Wenn der **Varianzwert** nicht angegeben ist, können Sie einen festen **g** angeben oder den Wert weglassen, um ihn zu integrieren.
5. Optional können Sie auf **Kriterien** klicken, um Einstellungen für „Bayes-Inferenz bei einer Stichprobe: Kriterien“ auf Seite 114“ (Zuverlässigkeitsintervall in Prozent, Optionen für fehlende Werte und

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

<sup>8</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Einstellungen für numerische Methoden) anzugeben, oder klicken Sie auf **A-priori-Verteilungen**, um Einstellungen für „Bayes-Inferenz bei einer Stichprobe: Binomiale/Poisson-A-priori-Wahrscheinlichkeiten“ auf Seite 118“ (konjugierte oder benutzerdefinierte A-priori-Verteilungen) anzugeben.

## Bayes-Inferenz bei unabhängigen Stichproben

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Die Prozedur "Bayes-Inferenz bei unabhängigen Stichproben" bietet Optionen für die Verwendung einer Gruppenvariablen, um zwei nicht zusammengehörige Gruppen zu definieren, und stellt die Bayes-Inferenz für die Differenz der beiden Gruppenmittelwerte bereit. Sie können die Bayes-Faktoren mithilfe verschiedener Methoden schätzen und außerdem die A-posteriori-Verteilung unter der Annahme charakterisieren, dass die Varianzen bekannt oder unbekannt sind.

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayes-Statistik > Unabhängige Stichproben, normal**

2. Wählen Sie in der Quellenvariablenliste die entsprechenden **Testvariablen** aus. Es muss mindestens eine Variable ausgewählt werden.

3. Wählen Sie in der Liste **Verfügbare Variablen** die entsprechende Gruppierungsvariable aus. Eine Gruppierungsvariable definiert zwei Gruppen für den unpaarigen *t*-Test. Die ausgewählte Gruppierungsvariable kann eine numerische Variable oder eine Zeichenfolgevariable sein.

4. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 6. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

9

10

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.
5. Verwenden Sie die Optionen für **Gruppen definieren**, um zwei Gruppen für den  $t$ -Test zu definieren, indem Sie zwei Werte (für Zeichenfolgevariablen) bzw. zwei Werte, einen Mittelpunkt oder einen Trennwert (für numerische Variablen) angeben.

**Anmerkung:** Die angegebenen Werte müssen in der Variablen vorhanden sein, andernfalls wird eine Fehlermeldung angezeigt, in der Sie darüber informiert werden, dass mindestens eine der Gruppen leer ist.

Für numerische Variablen:

- **Angegebene Werte verwenden.** Geben Sie einen Wert für Gruppe 1 und einen anderen Wert für Gruppe 2 ein. Fälle mit anderen Werten werden von der Analyse ausgeschlossen. Zahlen müssen nicht ganzzahlig sein (so sind beispielsweise 6,25 und 12,5 gültige Werte).
- **Mittelpunktwert verwenden.** Wenn diese Option ausgewählt ist, werden die Gruppen in Werte aufgeteilt, die kleiner als und größer-gleich dem Mittelpunkt sind.
- **Trennwert verwenden.**
  - **Trennwert.** Geben Sie eine Zahl ein, welche die Werte der Gruppierungsvariablen in zwei Mengen aufteilt. Alle Fälle mit Werten, die kleiner als der Trennwert sind, bilden eine Gruppe. Die Fälle mit Werten größer-gleich dem Trennwert bilden die andere Gruppe.

Für Zeichenfolge-Gruppierungsvariablen geben Sie eine Zeichenfolge für Gruppe 1 und einen anderen Wert für Gruppe 2 ein, z. B. *Ja* und *Nein*. Fälle mit anderen Zeichenfolge werden von der Analyse ausgeschlossen.

6. Sie können optional auf **Kriterien** klicken, um Einstellungen für "„Bayes-Inferenz bei unabhängigen Stichproben: Kriterien“ auf Seite 124" (Zuverlässigkeitsintervall in Prozent, Optionen für fehlende Werte und Einstellungen für die adaptive Quadraturmethode) anzugeben, oder klicken Sie auf **A-priori-Wahrscheinlichkeiten**, um Einstellungen für "„Bayes-Inferenz bei unabhängigen Stichproben: A-priori-Verteilung“ auf Seite 124" (Datenvarianz, A-priori-Wahrscheinlichkeit für Varianz und A-priori-Verteilung für bedingten Mittelwert für Varianz) anzugeben, oder Sie klicken auf **Bayes-Faktor schätzen**, um Einstellungen für "„Bayes-Inferenz bei unabhängigen Stichproben: Bayes-Faktor schätzen“ auf Seite 125" anzugeben.

## Bayes-Inferenz bei unabhängigen Stichproben: Gruppen definieren (numerisch)

Definieren Sie bei numerischen Gruppierungsvariablen die zwei Gruppen für den  $t$ -Test, indem Sie zwei Werte, einen Mittelpunkt oder einen Trennwert angeben.

**Anmerkung:** Die angegebenen Werte müssen in der Variablen vorhanden sein, andernfalls wird eine Fehlermeldung angezeigt, in der Sie darüber informiert werden, dass mindestens eine der Gruppen leer ist.

- **Angegebene Werte verwenden.** Geben Sie einen Wert für Gruppe 1 und einen anderen Wert für Gruppe 2 ein. Fälle mit anderen Werten werden von der Analyse ausgeschlossen. Zahlen müssen nicht ganzzahlig sein (so sind beispielsweise 6,25 und 12,5 gültige Werte).
- **Mittelpunktwert verwenden.** Wenn diese Option ausgewählt ist, werden die Gruppen in Werte aufgeteilt, die kleiner als und größer-gleich dem Mittelpunkt sind.
- **Trennwert verwenden.**
  - **Trennwert.** Geben Sie eine Zahl ein, welche die Werte der Gruppierungsvariablen in zwei Mengen aufteilt. Alle Fälle mit Werten, die kleiner als der Trennwert sind, bilden eine Gruppe. Die Fälle mit Werten größer-gleich dem Trennwert bilden die andere Gruppe.

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

<sup>10</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

## Bayes-Inferenz bei unabhängigen Stichproben: Gruppen definieren (Zeichenfolge)

Für Zeichenfolge-Gruppierungsvariablen geben Sie eine Zeichenfolge für Gruppe 1 und einen anderen Wert für Gruppe 2 ein, z. B. *Ja* und *Nein*. Fälle mit anderen Zeichenfolge werden von der Analyse ausgeschlossen.

**Anmerkung:** Die angegebenen Werte müssen in der Variablen vorhanden sein, andernfalls wird eine Fehlermeldung angezeigt, in der Sie darüber informiert werden, dass mindestens eine der Gruppen leer ist.

## Bayes-Inferenz bei unabhängigen Stichproben: Kriterien

Sie können die folgenden Analyse Kriterien für die Bayes-Inferenz bei unabhängigen Stichproben angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

### Fehlende Werte

Geben Sie die Methode zum Handhaben von fehlenden Werte an.

#### Paarweiser Fallausschluss

Dies ist die Standardeinstellung. Sie schließt Datensätze mit fehlenden Werten auf einer Analyse-für-Analyse-Basis aus. Datensätze mit fehlenden Werten für ein Feld, das für einen bestimmten Test verwendet wird, werden aus diesem Test ausgeschlossen.

#### Fälle listenweise ausschließen

Diese Einstellung schließt Datensätze mit fehlenden Werten listenweise aus. Datensätze mit fehlenden Werten für ein beliebiges Feld, das in einem beliebigen Unterbefehl genannt wird, werden aus allen Analysen ausgeschlossen.

**Anmerkung:** Die folgenden Optionen sind nur verfügbar, wenn die Option **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** für **Bayes-Analyse** ausgewählt ist.

### Adaptive Quadraturmethode

Geben Sie die Werte für die Toleranz und die maximale Anzahl der Iterationen für die adaptive Quadraturmethode an.

#### Toleranz

Geben Sie den Toleranzwert für die numerischen Methoden an. Die Standardeinstellung ist 0,000001.

#### Maximalzahl der Iterationen

Geben Sie die maximale Anzahl der Iterationen der adaptiven Quadraturmethode an. Dieser Wert muss eine positive Ganzzahl sein. Die Standardeinstellung ist 500.

## Bayes-Inferenz bei unabhängigen Stichproben: A-priori-Verteilung

Sie können die folgenden A-priori-Verteilungskriterien für die Bayes-Inferenz bei unabhängigen Stichproben angeben:

**Anmerkung:** Viele Anwendungsforscher stellen möglicherweise die Notwendigkeit zum Angeben einer A-priori-Wahrscheinlichkeit in Frage. A-priori-Wahrscheinlichkeiten als Referenz minimieren das Problem, wenn die A-priori-Wahrscheinlichkeit bei zunehmenden Daten überlastet ist. Wenn A-priori-Wahrscheinlichkeiten als Informationen angegeben werden, können Bayes-Methoden diese Informationen effektiv nutzen. Die Anforderung zum Angeben einer A-priori-Wahrscheinlichkeit sollte nicht als Hinderungsgrund für die Verwendung einer Bayes-Analyse betrachtet werden.

### Datenvarianz

Stellt Optionen zum Definieren von Einstellungen für die Datenvarianz bereit.

### **Varianz bekannt**

Wenn diese Option ausgewählt ist, können Sie zwei bekannte Gruppenvarianzen eingeben. Beide Werte müssen  $> 0$  sein.

#### **Varianz von Gruppe 1**

Geben Sie den ersten Wert für die bekannte Gruppenvarianz ein.

#### **Varianz von Gruppe 2**

Geben Sie den zweiten Wert für die bekannte Gruppenvarianz ein.

### **Varianzgleichheit voraussetzen**

Steuert, ob angenommen wird, dass die beiden Gruppenvarianzen gleich sind. Standardmäßig wird von ungleichen Gruppenvarianzen ausgegangen. Diese Einstellung wird ignoriert, wenn Werte für die beiden Gruppenvarianzen eingegeben werden.

### **Varianzungleichheit voraussetzen**

Steuert, ob angenommen wird, dass die beiden Gruppenvarianzen ungleich sind. Standardmäßig wird von ungleichen Gruppenvarianzen ausgegangen. Diese Einstellung wird ignoriert, wenn Werte für die beiden Gruppenvarianzen eingegeben werden.

### **A-priori-Verteilung für Varianz**

Geben Sie die A-priori-Verteilung für die beiden gleichen Varianzen an.

#### **Jeffreys**

Wenn diese Option ausgewählt ist, wird eine nicht informative A-priori-Verteilung (Ziel) für einen Parameterbereich verwendet.

#### **Invers Chi-Quadrat**

Gibt die stetige Wahrscheinlichkeitsverteilung einer Zufallsvariablen mit positiven Werten und die Parameter für inverse- $\chi^2(v_0, \sigma^2_0)$  an, wobei  $v_0 > 0$  der Freiheitsgrad und  $\sigma^2_0 > 0$  der Skalenparameter ist.

#### **Freiheitsgrade**

Geben Sie einen Wert für die Anzahl der Werte in der endgültigen Berechnung an, die frei variieren können.

#### **Skalenparameter**

Geben Sie den Skalenparameter  $\sigma^2_0 > 0$  für inverse- $\chi^2(v_0, \sigma^2_0)$  an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist. Je größer der Skalenparameter ist, desto mehr verteilt sich die Verteilung.

### **A-priori-Verteilung für bedingten Mittelwert für Varianz**

Stellt Optionen zum Angeben der A-priori-Verteilung für die beiden Gruppenmittelwerte bereit.

**Anmerkung:** Die Optionen **Diffus** und **Normal** sind nur verfügbar, wenn die Option **Varianz bekannt** ausgewählt ist.

#### **Diffus**

Die Standardeinstellung. Gibt die diffuse A-priori-Verteilung an.

#### **Normal**

Wenn diese Option ausgewählt ist, müssen Sie Lage- und Skalenparameter für die definierten Gruppenmittelwerte angeben.

#### **Lageparameter**

Geben Sie numerische Werte ein, die die Lageparameter für die Gruppenverteilungen angeben.

#### **Skalenparameter**

Geben Sie den Skalenparameter  $\sigma^2_0 > 0$  für inverse- $\chi^2(v_0, \sigma^2_0)$  an. Für jede Gruppe müssen Sie einen einzelnen Wert eingeben, der größer als 0 ist. Je größer der Skalenparameter ist, desto mehr verteilt sich die Verteilung.

## **Bayes-Inferenz bei unabhängigen Stichproben: Bayes-Faktor schätzen**

Sie können die Methode zum Schätzen des Bayes-Faktors angeben.

### Rouder-Methode

Wenn diese Option ausgewählt ist, wird die Rouder-Methode aufgerufen. Dies ist die Standardeinstellung.

### Gonen-Methode

Wenn diese Option ausgewählt ist, wird die Gonen-Methode aufgerufen und Sie müssen die folgenden Einstellungen für die Effektgröße angeben:

#### Mittelwert für Effektgröße

Geben Sie einen Wert ein, der die Mittelwertdifferenz zwischen den beiden Gruppen angibt.

#### Varianz für Effektgröße

Geben Sie einen Wert ein, der die Varianz für die beiden Gruppen angibt. Der Wert muss > 0 sein.

### Hyper-A-priori-Methode

Wenn diese Option ausgewählt ist, wird die Hyper-g-Methode aufgerufen, bei der Sie einen einzelnen Wert angeben müssen. Geben Sie einen Wert zwischen -1 und -0,5 in das Feld **Formparameter** ein. Der Standardwert ist -0,75.

## Bayessche Inferenz über Pearson-Korrelation

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Der Pearson-Korrelationskoeffizient misst die lineare Beziehung zwischen zwei metrischen Variablen, die gemeinsam einer bivariaten Normalverteilung folgen. Die konventionelle statistische Inferenz über Korrelationskoeffizienten wurde umfassend diskutiert und das Verfahren wird seit langem in IBM SPSS Statistics angeboten. Mit dem Design der bayesschen Inferenz über Pearson-Korrelationskoeffizienten können Benutzer die Bayes-Inferenz zeichnen, indem sie Bayes-Faktoren schätzen und A-posteriori-Verteilungen charakterisieren.

1. Wählen Sie in den Menüs Folgendes aus:

#### **Analysieren > Bayessche Statistik > Pearson-Korrelation**

2. Wählen Sie in der Liste **Verfügbare Variablen** die entsprechenden Testvariablen aus, die für die paarweise Korrelationsinferenz verwendet werden sollen. Es müssen mindestens zwei Quellenvariablen angegeben werden. Wenn mehr als zwei Variablen ausgewählt sind, wird die Analyse für alle paarweisen Kombinationen der ausgewählten Variablen ausgeführt.

3. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0

Tabelle 7. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen (Forts.)

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

11

12

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.
4. Geben Sie die maximale Anzahl Diagramme an, die in der Ausgabe angezeigt werden sollen. Ein Set von Diagrammen kann drei Diagramme in demselben Fensterbereich enthalten. Die Diagramme werden in der Reihenfolge von der ersten Variablen im Vergleich zu den übrigen Variablen erzeugt, dann der zweiten Variablen im Vergleich zu den übrigen Variablen und so weiter. Der definierte ganzzahlige Wert muss zwischen 0 und 50 liegen. Standardmäßig werden 10 Sets von Diagrammen für fünf Variablen ausgegeben. Diese Option ist nicht verfügbar, wenn **Bayes-Faktor schätzen** ausgewählt ist.
  5. Optional können Sie auf **Kriterien** klicken, um Einstellungen für „Bayessche Pearson-Korrelation: Kriterien“ auf Seite 127“ (Zuverlässigkeitsintervall in Prozent, Optionen für fehlende Werte und Einstellungen für numerische Methoden) anzugeben, oder klicken Sie auf **A-priori-Wahrscheinlichkeiten**, um Einstellungen für „Bayessche Pearson-Korrelation: A-priori-Verteilung“ auf Seite 128“ (Wert  $c$  für die A-priori-Verteilung  $p(\rho) \propto (1 - \rho^2)^c$ ) anzugeben, oder klicken Sie auf **Bayes-Faktor**, um Einstellungen für „Bayes-Inferenz bei unabhängigen Stichproben: Bayes-Faktor schätzen“ auf Seite 125“ anzugeben.

## Bayessche Pearson-Korrelation: Kriterien

Sie können die folgenden Analyse Kriterien für die bayessche Pearson-Korrelationsinferenz angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

### Fehlende Werte

Geben Sie die Methode zum Handhaben von fehlenden Werte an.

#### Paarweiser Fallausschluss

Diese Einstellung schließt Datensätze mit fehlenden Werten paarweise aus.

#### Fälle listenweise ausschließen

Diese Einstellung schließt Datensätze mit fehlenden Werten listenweise aus. Datensätze mit fehlenden Werten für ein beliebiges Feld, das in einem beliebigen Unterbefehl genannt wird, werden aus allen Analysen ausgeschlossen.

**Anmerkung:** Die folgenden Optionen sind nur verfügbar, wenn die Option **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** für **Bayes-Analyse** ausgewählt ist.

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

<sup>12</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

## Numerische Methode

Geben Sie die numerische Methode an, die zum Schätzen des Integrals verwendet wird.

### Benutzerdefinierten Startwert festlegen

Bei Auswahl dieser Option können Sie im Feld **Start** einen benutzerdefinierten Startwert angeben.

### Toleranz

Geben Sie den Toleranzwert für die numerischen Methoden an. Die Standardeinstellung ist 0,000001.

### Maximalzahl der Iterationen

Geben Sie die maximale Anzahl der Methodeniterationen an. Dieser Wert muss eine positive Ganzzahl sein. Die Standardeinstellung ist 2000.

### Anzahl der Monte-Carlo-Stichproben

Geben Sie die Anzahl der Punkte an, für die Stichproben für die Monte-Carlo-Approximation gezogen werden sollen. Dieser Wert muss eine positive Ganzzahl sein. Der Standardwert ist 10000.

### Für A-posteriori-Verteilung simulierte Stichproben

Geben Sie die Anzahl der Stichproben an, die zum Zeichnen der gewünschten A-posteriori-Verteilung verwendet werden. Der Standardwert ist 10000.

## Bayessche Pearson-Korrelation: A-priori-Verteilung

Sie können den Wert  $c$  für die A-priori-Verteilung  $p(\rho) \propto (1-\rho^2)^c$  angeben.

**Anmerkung:** Viele Anwendungsforscher stellen möglicherweise die Notwendigkeit zum Angeben einer A-priori-Wahrscheinlichkeit in Frage. A-priori-Wahrscheinlichkeiten als Referenz minimieren das Problem, wenn die A-priori-Wahrscheinlichkeit bei zunehmenden Daten überlastet ist. Wenn A-priori-Wahrscheinlichkeiten als Informationen angegeben werden, können Bayes-Methoden diese Informationen effektiv nutzen. Die Anforderung zum Angeben einer A-priori-Wahrscheinlichkeit sollte nicht als Hinderungsgrund für die Verwendung einer Bayes-Analyse betrachtet werden.

### Gleichmäßig ( $c = 0$ )

Wenn diese Option ausgewählt ist, wird eine gleichmäßige A-priori-Verteilung verwendet.

### Jeffreys ( $c = -1,5$ )

Wenn diese Option ausgewählt ist, wird eine nicht informative A-priori-Verteilung verwendet.

### Benutzerdefinierten $c$ -Wert festlegen

Wenn diese Option ausgewählt ist, können Sie einen benutzerdefinierten  $c$ -Wert angeben. Jede einzelne reelle Zahl ist zulässig.

## Bayessche Pearson-Korrelation: Bayes-Faktor

Sie können die Methode zum Schätzen des Bayes-Faktors angeben. Die folgenden Optionen sind nur verfügbar, wenn entweder die Option **Bayes-Faktor schätzen** oder die Option **Beide Methoden verwenden** Bayes-Analyse ausgewählt ist.

### JZS-Bayes-Faktor

Wenn diese Option ausgewählt ist, wird die Zellner-Siow-Methode aufgerufen. Hierbei handelt es sich um die Standardeinstellung.

### Bruchzahl-Bayes-Faktor

Wenn diese Option ausgewählt ist, können Sie den Bruchzahl-Bayes-Faktor und den Nullhypothese-nwert angeben. Für den Bruchzahl-Bayes-Faktor müssen Sie einen Wert  $\epsilon$  (0,1) angeben. Der Standardwert ist 0,5.

## Bayessche Inferenz über lineare Regressionsmodelle

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Regression ist eine statistische Methode, die im Wesentlichen bei der quantitativen Modellierung verwendet wird. Lineare Regression ist ein grundlegendes Standardkonzept, bei dem Wissenschaftler Werte ei-

nes Skalenergebnisses mithilfe der Werte von mehreren Variablen erläutern oder vorhersagen. Bayessche univariate lineare Regression ist ein Konzept für lineare Regression, bei dem die statistische Analyse im Rahmen der bayesschen Inferenz durchgeführt wird.

Sie können die Regressionsprozedur aufrufen und ein vollständiges Modell definieren.

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayes-Statistik > Lineare Regression**

2. Wählen Sie eine einzelne abhängige Variable aus der Liste **Verfügbare Variablen** aus, die keine Zeichenfolgevariable ist. Sie müssen eine Variable auswählen, die keine Zeichenfolgevariable ist.
3. Wählen Sie mindestens eine kategoriale Faktorvariable für das Modell aus der Liste **Verfügbare Variablen** aus.
4. Wählen Sie mindestens eine metrische Kovariatenvariable aus der Liste **Verfügbare Variablen** aus, die keine Zeichenfolgevariable ist.

**Anmerkung:** Die beiden Listen **Faktor(en)** und **Kovariate(n)** dürfen nicht leer sein. Sie müssen mindestens eine Variable für **Faktor(en)** oder **Kovariate(n)** auswählen.

5. Wählen Sie optional für die Regressionsgewichtung eine einzelne Variable aus der Liste **Verfügbare Variablen** aus, die keine Zeichenfolgevariable ist.
6. Wählen Sie die gewünschte **Bayes-Analyse** aus:
  - **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
  - **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 8. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

13

14

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

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.

Darüber hinaus haben Sie die folgenden Möglichkeiten:

- Klicken Sie auf **Kriterien**, um die Einstellungen für das Zuverlässigkeitsintervall in Prozent und die numerische Methode anzugeben.
- Klicken Sie auf **A-priori-Verteilungen**, um die Einstellungen für Referenz- und konjugierte A-priori-Verteilungen zu definieren.
- Klicken Sie auf **Bayes-Faktor**, um Einstellungen für Bayes-Faktoren anzugeben.
- Klicken Sie auf **Speichern**, um anzugeben, welche Elemente gespeichert werden sollen, und speichern Sie Modellinformationen in einer XML-Datei.
- Klicken Sie auf **Vorhersage**, um Regressoren für die Bayes-Vorhersage anzugeben.
- Klicken Sie auf **Diagramme**, um die A-posteriori-Verteilungen der Regressionsparameter, die Varianz der Fehlerterme und die vorhergesagten Werte abzubilden.
- Klicken Sie auf **F-Tests**, um durch einen Vergleich der statistischen Modelle das Modell zu identifizieren, das am besten zur Grundgesamtheit passt, aus der es erfasst wurde.

## Bayessche lineare Regressionsmodelle: Kriterien

Sie können die folgenden Analyse Kriterien für bayessche lineare Regressionsmodelle angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

**Anmerkung:** Die folgenden Optionen sind nur verfügbar, wenn die Option **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** für **Bayes-Analyse** ausgewählt ist.

### Numerische Methode

Geben Sie die numerische Methode an, die zum Schätzen des Integrals verwendet wird.

### Toleranz

Geben Sie den Toleranzwert für die numerischen Methoden an. Die Standardeinstellung ist 0,000001.

### Maximalzahl der Iterationen

Geben Sie die maximale Anzahl der Methodeniterationen an. Dieser Wert muss eine positive Ganzzahl sein. Die Standardeinstellung ist 2000.

## Bayessche lineare Regressionsmodelle: A-priori-Verteilungen

Sie können die folgenden Einstellungen für A-priori-Verteilungen für die Regressionsparameter und die Varianz der Fehler angeben. Die folgenden Optionen sind nur verfügbar, wenn die Option **A-posteriori-Verteilung charakterisieren** für **Bayes-Analyse** ausgewählt ist.

**Anmerkung:** Viele Anwendungsforscher stellen möglicherweise die Notwendigkeit zum Angeben einer A-priori-Wahrscheinlichkeit in Frage. A-priori-Wahrscheinlichkeiten als Referenz minimieren das Problem, wenn die A-priori-Wahrscheinlichkeit bei zunehmenden Daten überlastet ist. Wenn A-priori-Wahrscheinlichkeiten als Informationen angegeben werden, können Bayes-Methoden diese Informationen effektiv nutzen. Die Anforderung zum Angeben einer A-priori-Wahrscheinlichkeit sollte nicht als Hinderungsgrund für die Verwendung einer Bayes-Analyse betrachtet werden.

### Referenz

Wenn diese Option ausgewählt ist, erstellt die Referenzanalyse eine objektive bayessche Inferenz. Inferenzanweisungen hängen nur vom angenommenen Modell und den verfügbaren Daten ab. Die A-priori-Verteilung, die zum Erstellen einer Inferenz verwendet wird, ist am wenigsten informativ. Hierbei handelt es sich um die Standardeinstellung.

<sup>14</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

## Konjugieren

Stellt Optionen zum Definieren von konjugierten A-priori-Verteilungen bereit. Konjugierte A-priori-Verteilungen setzen die normale, inverse, gemeinsame Gammaverteilung voraus. Konjugierte A-priori-Verteilungen sind zwar für die Ausführung von Bayes-Aktualisierungen nicht erforderlich, aber sie sind bei den Berechnungsprozessen hilfreich.

**Anmerkung:** Legen Sie in der Tabelle **A-priori-Verteilung für Varianz von Fehlern** den erwarteten Mittelwert von Regressionsparametern fest, um konjugierte A-priori-Wahrscheinlichkeiten für ein lineares Regressionsmodell anzugeben. Sie können auch die Einstellung **Varianz der Kovarianzmatrix** verwenden, um die Varianz/Kovarianz der A-priori-Wahrscheinlichkeit anzugeben.

### A-priori-Verteilung für Varianz von Fehlern

#### Formparameter

Geben Sie den Formparameter  $a_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist.

#### Skalenparameter

Geben Sie den Skalenparameter  $b_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist. Je größer der Skalenparameter ist, desto mehr verteilt sich die Verteilung.

In der Tabelle wird der Mittelwert von Regressionsparametern (einschließlich konstantem Term) aufgelistet, der den Mittelwertvektor  $\vartheta_0$  für die definierten Regressionsparameter angibt. Die Anzahl der Werte muss der Anzahl der Regressionsparameter entsprechen, einschließlich der konstanten Terme.

Der erste Variablenname ist immer INTERCEPT. Aus der zweiten Zeile wird die Spalte **Variablen** automatisch mit den Variablen gefüllt, die durch Faktor (e) und Kovariate(n) angegeben werden. Die Spalte **Mittelwert** enthält keine Standardwerte.

Klicken Sie auf **Zurücksetzen**, um die Werte zu löschen.

### Varianz der Kovarianzmatrix: $\sigma^2\mathbf{x}$

Geben Sie die  $V_0$ -Werte im oberen Dreieck der Varianz-Kovarianz-Matrix für die multivariate normale A-priori-Verteilung an. Beachten Sie, dass  $V_0$  semipositiv definit sein muss. Der letzte Wert in jeder Zeile muss positiv sein. Die nächste Zeile sollte einen Wert mehr als die vorherige Zeile aufweisen. Für Referenzkategorien (sofern vorhanden) werden keine Werte angegeben.

Klicken Sie auf **Zurücksetzen**, um die Werte zu löschen.

### Identitätsmatrix verwenden

Wenn diese Option ausgewählt ist, wird die skalierte Identitätsmatrix verwendet. Sie können im unteren Dreieck der Varianz-Kovarianz-Matrix für die multivariate normale A-priori-Verteilung keine  $V_0$ -Werte angeben.

## Bayessche lineare Regressionsmodelle: Bayes-Faktor

Sie können das Modelldesign für die Analyse angeben, inklusive der Methode zum Schätzen des Bayes-Faktors für Modelle mit bayesscher linearer Regression. Die folgenden Optionen sind nur verfügbar, wenn entweder die Option **Bayes-Faktor schätzen** oder die Option **Beide Methoden verwenden** Bayesian Analysis ausgewählt ist.

### Nullmodell

Wenn diese Option ausgewählt ist, basieren die geschätzten Bayes-Faktoren auf dem Nullmodell. Hierbei handelt es sich um die Standardeinstellung.

### Gesamtes Modell

Wenn diese Option ausgewählt ist, basieren die geschätzten Bayes-Faktoren auf dem gesamten Modell und Sie können zu verwendende Variablen und weitere Faktoren und Kovariaten auswählen.

### Variablen

Listet alle Variablen auf, die für das gesamte Modell verfügbar sind.

### **Weitere(r) Faktor(en)**

Wählen Sie Variablen aus der Variablenliste aus, die als weitere Faktoren verwendet werden sollen.

### **Weitere Kovariate(n)**

Wählen Sie Variablen aus der Variablenliste aus, die als weitere Kovariaten verwendet werden sollen.

### **Berechnung**

Geben Sie die Methode zum Schätzen von Bayes-Faktoren an. Die JZS-Methode ist die Standardeinstellung.

#### **JZS-Methode**

Wenn diese Option ausgewählt ist, wird die Zellner-Siow-Methode aufgerufen. Hierbei handelt es sich um die Standardeinstellung.

#### **Zellner-Methode**

Wenn diese Option ausgewählt ist, wird die Zellner-Methode aufgerufen und Sie müssen einen einzelnen A-priori-Wert  $g > 0$  eingeben. (Es gibt keinen Standardwert.)

#### **Hyper-A-priori-Methode**

Wenn diese Option ausgewählt ist, wird die Hyper- $g$ -Methode aufgerufen und Sie müssen einen Formparameter  $\alpha_0$  für inverse Gammaverteilung angeben. Sie müssen einen einzelnen Wert  $> 0$  angeben. (Der Standardwert ist 3.)

#### **Rouder-Methode**

Wenn diese Option ausgewählt ist, wird die Rouder-Methode aufgerufen und Sie müssen einen Skalenparameter  $b_0$  für inverse Gammaverteilung angeben. Sie müssen einen einzelnen Wert  $> 0$  angeben (der Standardwert ist 1).

## **Bayessche lineare Regressionsmodelle: Speichern**

In diesem Dialogfeld können Sie angeben, welche Statistiken für die Bayes-Vorhersageverteilung bewertet werden, und die Modellergebnisse in eine XML-Datei exportieren.

### **A-Posteriori-Vorhersagestatistik**

Sie können die folgenden Statistiken bewerten, die von Bayes-Vorhersagen abgeleitet werden.

#### **Mittelwert**

Mittelwert der A-posteriori-Vorhersageverteilung.

#### **Varianzen**

Varianz der A-posteriori-Vorhersageverteilung.

#### **Modi**

Modus der A-posteriori-Vorhersageverteilung.

#### **Untergrenze für Zuverlässigkeitsintervall**

Untergrenze des Zuverlässigkeitsintervalls der A-posteriori-Vorhersageverteilung.

#### **Obergrenze für Zuverlässigkeitsintervall**

Obergrenze des Zuverlässigkeitsintervalls der A-posteriori-Vorhersageverteilung.

**Anmerkung:** Sie können für jede Statistik entsprechende Variablennamen zuordnen.

### **Modellinformationen in XML-Datei exportieren**

Geben Sie den Namen und den Speicherort einer XML-Datei an, um die Varianz-Kovarianz-Matrix der bewerteten Parameter zu exportieren.

## **Bayessche lineare Regressionsmodelle: Vorhersagen**

Sie können die Regressoren angeben, um Vorhersageverteilungen zu generieren.

### **Regressoren für Bayes-Vorhersage**

Die Tabelle enthält alle verfügbaren Regressoren. Die Spalte **Regressoren** wird automatisch mit bestimmten Faktor- und Kovariatenvariablen ausgefüllt. Geben Sie die beobachteten Vektoren mit den Werten für die Regressoren an. Jedem Regressor kann ein Wert oder eine Zeichenfolge zugeordnet

werden und er darf nur zum Vorhersagen eines Falls verwendet werden. Für Faktoren sind sowohl Werte als auch Zeichenfolgen zulässig.

Um die Vorhersage (durch Klicken auf **Analyse ausführen**) ausführen zu können, müssen Sie entweder alle oder keine Regressorwerte angeben.

Wenn eine Faktor- oder Kovariatenvariable entfernt wird, wird die entsprechende Regressorzeile aus der Tabelle entfernt.

Für Kovariaten können nur numerische Werte angegeben werden. Für Faktoren sind sowohl numerische Werte als auch Zeichenfolgen zulässig.

**Anmerkung:** Klicken Sie auf **Zurücksetzen**, um die definierten Werte zu löschen.

## Bayessche lineare Regressionsmodelle: Diagramme

Sie können die ausgegebenen Diagramme steuern.

### Kovariaten

Listet die zurzeit definierten Kovariaten auf.

### Kovariaten darstellen

Wählen Sie die Kovariaten, die dargestellt werden sollen, aus der Liste **Kovariate(n)** aus und fügen Sie sie der Liste **Kovariaten darstellen** hinzu.

### Faktoren

Listet die zurzeit definierten Faktoren auf.

### Faktor(en) darstellen

Wählen Sie die Faktoren, die dargestellt werden sollen, aus der Liste **Faktor(en)** aus und fügen Sie sie der Liste **Faktor(en) darstellen** hinzu.

### Maximal darzustellende Kategorien

Wählen Sie die maximale Anzahl der Kategorien aus, die dargestellt werden sollen (einzelne, positive ganze Zahl). Diese Einstellung gilt für alle Faktoren. Standardmäßig werden die ersten 2 Stufen für jeden Faktor dargestellt.

### Diagramme einschließen von

#### Konstanter Term

Wenn diese Option ausgewählt ist, wird der konstante Term dargestellt. Diese Einstellung ist standardmäßig inaktiviert.

#### Varianz der Fehlerterme

Wenn diese Option ausgewählt ist, wird die Varianz der Fehler dargestellt. Diese Einstellung ist standardmäßig inaktiviert.

#### Vorhergesagte Bayes-Verteilung

Wenn diese Option ausgewählt ist, wird die Vorhersageverteilung dargestellt. Diese Einstellung ist standardmäßig inaktiviert. Die Einstellung kann nur ausgewählt werden, wenn gültige Regressorwerte ausgewählt wurden.

## Bayessche lineare Regressionsmodelle: F-Tests

Sie können einen oder mehrere partielle F-Tests erstellen. Ein F-Test ist jeder statistische Test, bei dem die Teststatistik eine F-Verteilung unter der Nullhypothese hat. F-Tests werden häufig verwendet, wenn statistische Modelle verglichen werden, die an ein Dataset angepasst wurden, um das Modell zu identifizieren, das am besten zur Grundgesamtheit passt, aus der Stichprobendaten erfasst wurden.

### Verfügbare Variablen

Listet die Faktor- und Kovariatenvariablen auf, die im Hauptdialogfeld für die bayessche lineare Regression ausgewählt werden. Wenn Faktor- und Kovariatenvariablen über das Hauptdialogfeld hinzugefügt oder entfernt werden, wird die Liste entsprechend aktualisiert.

### Testvariable(n)

Wählen Sie die Faktor-/Kovariatenvariablen, die getestet werden sollen, aus der Liste **Verfügbare Variablen** aus und fügen Sie sie der Liste **Testvariable(n)** hinzu.

**Anmerkung:** Die Option **Konstanten Term einschließen** muss ausgewählt werden, wenn keine Testfaktoren oder -kovariaten ausgewählt sind.

**Testvariable(n) und -wert(e)**

Geben Sie die Werte an, die getestet werden sollen. Die Anzahl der Werte muss der Anzahl der Parameter im ursprünglichen Modell entsprechen. Wenn Werte angegeben werden, muss der erste Wert für den konstanten Term angegeben werden (es wird vorausgesetzt, dass alle Werte 0 sind, wenn sie nicht explizit definiert sind).

**Konstanten Term einschließen**

Wenn diese Option ausgewählt ist, werden die konstanten Terme in den Test eingeschlossen. Diese Einstellung ist standardmäßig nicht ausgewählt.

Wenn diese Option aktiviert ist, verwenden Sie das Feld **Testwert**, um einen Wert anzugeben.

**Testbeschriftung (optional)**

Optional können Sie eine Beschriftung für jeden Test angeben. Sie können einen Zeichenfolgewart mit einer maximalen Länge von 255 Byte angeben. Pro F-Test ist jeweils nur eine Beschriftung zulässig.

## Bayessche einfaktorielle ANOVA

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Die Prozedur "Einfaktorielle ANOVA" führt eine einfaktorielle Varianzanalyse für eine quantitative abhängige Variable mit einer einzelnen (unabhängigen) Faktorvariablen durch. Die Analyse der Varianz wird verwendet, um die Hypothese zu testen, dass mehrere Mittel gleich sind. SPSS Statistics unterstützt Bayes-Faktoren, konjugierte und nicht informative A-priori-Wahrscheinlichkeiten.

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayes-Statistik > Einfaktorielle ANOVA**

2. Wählen Sie eine einzelne, numerische abhängige Variable aus der Liste **Verfügbare Variablen** aus. Sie müssen mindestens eine Variable auswählen.

3. Wählen Sie eine einzelne Faktorvariable für das Modell aus der Liste **Verfügbare Variablen** aus. Sie müssen mindestens eine **Faktorvariable** auswählen.

4. Wählen Sie für die Regressionsgewichtung eine einzelne Variable aus der Liste **Verfügbare Variablen** aus, die keine Zeichenfolgevariable ist. Das Variablenfeld **Gewichtung** kann leer sein.

5. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 9. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0

Tabelle 9. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen (Forts.)

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

$H_0$ : Nullhypothese

$H_1$ : Alternativhypothese

15

16

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.

Die folgenden Optionen sind verfügbar:

- Klicken Sie auf **Kriterien**, um die Einstellungen für das Zuverlässigkeitsintervall in Prozent und die numerische Methode anzugeben.
- Klicken Sie auf **A-priori-Verteilungen**, um die Einstellungen für Referenz- und konjugierte A-priori-Verteilungen zu definieren.
- Klicken Sie auf **Bayes-Faktor**, um Einstellungen für Bayes-Faktoren anzugeben.
- Klicken Sie auf **Diagramme**, um die ausgegebenen Diagramme zu steuern.

## Bayessche einfaktorielle ANOVA: Kriterien

Sie können die folgenden Analyse Kriterien für bayessche einfaktorielle ANOVA-Modelle angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

**Anmerkung:** Die folgenden Optionen sind nur verfügbar, wenn die Option **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** für **Bayes-Analyse** ausgewählt ist.

### Numerische Methode

Geben Sie die numerische Methode an, die zum Schätzen des Integrals verwendet wird.

### Toleranz

Geben Sie den Toleranzwert für die numerischen Methoden an. Die Standardeinstellung ist 0,000001.

### Maximalzahl der Iterationen

Geben Sie die maximale Anzahl der Methodeniterationen an. Dieser Wert muss eine positive Ganzzahl sein. Die Standardeinstellung ist 2000.

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

<sup>16</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

## Bayessche einfaktorielle ANOVA: A-priori-Wahrscheinlichkeiten

Sie können die folgenden Einstellungen für A-priori-Verteilungen für die Regressionsparameter und die Varianz der Fehler angeben. Die folgenden Optionen sind nur verfügbar, wenn die Option **A-posteriori-Verteilung charakterisieren** für **Bayes-Analyse** ausgewählt ist.

**Anmerkung:** Viele Anwendungsforscher stellen möglicherweise die Notwendigkeit zum Angeben einer A-priori-Wahrscheinlichkeit in Frage. A-priori-Wahrscheinlichkeiten als Referenz minimieren das Problem, wenn die A-priori-Wahrscheinlichkeit bei zunehmenden Daten überlastet ist. Wenn A-priori-Wahrscheinlichkeiten als Informationen angegeben werden, können Bayes-Methoden diese Informationen effektiv nutzen. Die Anforderung zum Angeben einer A-priori-Wahrscheinlichkeit sollte nicht als Hinderungsgrund für die Verwendung einer Bayes-Analyse betrachtet werden.

### Referenz

Wenn diese Option ausgewählt ist, erstellt die Referenzanalyse eine objektive bayessche Inferenz. Inferenzanweisungen hängen nur vom angenommenen Modell und den verfügbaren Daten ab. Die A-priori-Verteilung, die zum Erstellen einer Inferenz verwendet wird, ist am wenigsten informativ. Hierbei handelt es sich um die Standardeinstellung.

### Konjugieren

Stellt Optionen zum Definieren von konjugierten A-priori-Verteilungen bereit. Konjugierte A-priori-Verteilungen setzen die normale, inverse, gemeinsame Gammaverteilung voraus. Konjugierte A-priori-Verteilungen sind zwar für die Ausführung von Bayes-Aktualisierungen nicht erforderlich, aber sie sind bei den Berechnungsprozessen hilfreich.

### A-priori-Verteilung für Varianz von Fehlern

#### Formparameter

Geben Sie den Formparameter  $a_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist.

#### Skalenparameter

Geben Sie den Skalenparameter  $b_0$  für inverse Gammaverteilung an. Sie müssen einen einzelnen Wert eingeben, der größer als 0 ist. Je größer der Skalenparameter ist, desto mehr verteilt sich die Verteilung.

### A-priori-Verteilungen für Regressionsparameter

Geben Sie den Mittelwertvektor  $\beta_0$  für die Gruppenmittelwerte an. Die Anzahl der Werte muss der Anzahl der Regressionsparameter entsprechen, einschließlich der konstanten Terme.

Die Spalte **Variablen** wird automatisch mit den Stufen des Faktors ausgefüllt. Die Spalte **Mittelwert** enthält keine Standardwerte.

Klicken Sie auf **Zurücksetzen**, um die Werte zu löschen.

### Varianz der Kovarianzmatrix: $\sigma^2x$

Geben Sie die  $V_0$ -Werte im oberen Dreieck der Varianz-Kovarianz-Matrix für die multivariate normale A-priori-Verteilung an. Beachten Sie, dass  $V_0$  semipositiv definit sein muss. Nur das untere Dreieck der Tabelle muss angegeben werden.

Die Zeilen und Spalten werden automatisch mit den Stufen des Faktors ausgefüllt. Alle Werte in der Diagonalen sind 1; alle außerhalb der Diagonalen liegenden Werte sind 0.

Klicken Sie auf **Zurücksetzen**, um die Werte zu löschen.

### Identitätsmatrix verwenden

Wenn diese Option ausgewählt ist, wird die Identitätsmatrix verwendet. Sie können im unteren Dreieck der Varianz-Kovarianz-Matrix für die multivariate normale A-priori-Verteilung keine  $V_0$ -Werte angeben.

## Bayessche einfaktorielle ANOVA: Bayes-Faktor

Sie können die Methode zum Schätzen des Bayes-Faktors für Modelle mit bayesschem einfaktoriellem ANOVA angeben. Die folgenden Optionen sind nur verfügbar, wenn entweder die Option **Bayes-Faktor schätzen** oder die Option **Beide Methoden verwenden** Bayesian Analysis ausgewählt ist.

## Berechnung

Geben Sie die Methode zum Schätzen von Bayes-Faktoren an. Die JZS-Methode ist die Standardeinstellung.

### JZS-Methode

Wenn diese Option ausgewählt ist, wird die Zellner-Siow-Methode aufgerufen. Hierbei handelt es sich um die Standardeinstellung.

### Zellner-Methode

Wenn diese Option ausgewählt ist, wird die Zellner-Methode aufgerufen und Sie müssen einen einzelnen A-priori-Wert  $g > 0$  eingeben. (Es gibt keinen Standardwert.)

### Hyper-A-priori-Methode

Wenn diese Option ausgewählt ist, wird die Hyper-g-Methode aufgerufen und Sie müssen einen Formparameter  $\alpha_0$  für inverse Gammaverteilung angeben. Sie müssen einen einzelnen Wert  $> 0$  angeben. (Der Standardwert ist 3.)

### Rouder-Methode

Wenn diese Option ausgewählt ist, wird die Rouder-Methode aufgerufen und Sie müssen einen Skalenparameter  $b_0$  für inverse Gammaverteilung angeben. Sie müssen einen einzelnen Wert  $> 0$  angeben (der Standardwert ist 1).

## Bayessche einfaktorielle ANOVA: Diagramme

Sie können die ausgegebenen Diagramme steuern.

### Gruppe(n) darstellen

Geben Sie die Untergruppen an, die dargestellt werden sollen. Stellen Sie die Wahrscheinlichkeit (A-priori und A-posteriori) für die Mittelwerte der angegebenen Gruppen dar. Die Liste **Gruppen** ist ein Subset der Kategorien der Faktorvariablen. Das Format sollte daher mit dem Datentyp und den tatsächlichen Werten des Faktors konsistent sein.

### Varianz der Fehlerterme

Wenn diese Option ausgewählt ist, wird die Varianz der Fehler dargestellt. Diese Einstellung ist standardmäßig inaktiviert. Diese Option ist nicht verfügbar, wenn **Bayes-Faktor schätzen** als Bayes-Analyse ausgewählt ist.

## Bayessche loglineare Modelle

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

Das Design für das Testen der Unabhängigkeit zweier Faktoren erfordert zwei kategoriale Variablen für die Erstellung einer Kontingenztafel und führt eine bayessche Inferenz für die Zeilen/Spaltenzuordnung durch. Sie können die Bayes-Faktoren unter Annahme verschiedener Modelle schätzen und die gewünschte A-posteriori-Verteilung durch Simulieren des simultanen Zuverlässigkeitsintervalls für Interaktionsterme charakterisieren.

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayes-Statistik > Loglineare Modelle**

2. Wählen Sie aus der Liste **Verfügbare Variablen** eine einzelne, nicht metrische Zeilenvariable aus. Sie müssen mindestens eine nicht metrische Variable auswählen.
3. Wählen Sie aus der Liste **Verfügbare Variablen** eine einzelne, nicht metrische Spaltenvariable aus. Sie müssen mindestens eine nicht metrische Variable auswählen.
4. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.

- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 10. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

17

18

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.

Die folgenden Optionen sind verfügbar:

- Klicken Sie auf **Kriterien**, um die Einstellungen für das Zuverlässigkeitsintervall in Prozent und die numerische Methode anzugeben.
- Klicken Sie auf **Bayes-Faktor**, um Einstellungen für Bayes-Faktoren anzugeben.
- Klicken Sie auf **Drucken**, um anzugeben, wie der Inhalt in den Ausgabetafeln angezeigt werden soll.

## Bayessche loglineare Modelle: Kriterien

Sie können die folgenden Analyseverfahren für bayessche loglineare Regressionsmodelle angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

### Numerische Methode

Geben Sie die numerische Methode an, die zum Schätzen des Integrals verwendet wird.

### Benutzerdefinierten Startwert festlegen

Bei Auswahl dieser Option können Sie im Feld **Start** einen benutzerdefinierten Startwert angeben. Geben Sie einen Startwert für Zufallszahlen an. Dieser Wert muss eine positive Ganzzahl sein. Standardmäßig wird ein Startwert für Zufallszahlen zugewiesen.

**Anmerkung:** Die folgenden Optionen sind nur verfügbar, wenn die Option **Bayes-Faktor schätzen** oder **Beide Methoden verwenden** für **Bayes-Analyse** ausgewählt ist.

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

<sup>18</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### Toleranz

Geben Sie den Toleranzwert für die numerischen Methoden an. Die Standardeinstellung ist 0,000001.

### Maximalzahl der Iterationen

Geben Sie die maximale Anzahl der Methodeniterationen an. Dieser Wert muss eine positive Ganzzahl sein. Die Standardeinstellung ist 2000.

### Stichproben simuliert für A-posteriori-Verteilung

Geben Sie die Anzahl der Stichproben an, die zum Zeichnen der gewünschten A-posteriori-Verteilung verwendet werden. Der Standardwert ist 10000.

### Format

Wählen Sie aus, ob Kategorien in aufsteigender oder absteigender Reihenfolge angezeigt werden. Aufsteigend ist die Standardeinstellung.

## Bayessche loglineare Modelle: Bayes-Faktor

Sie können das angenommene Modell für die beobachteten Daten angeben (Poisson, multinomial oder nicht parametrisch). Die multinomiale Verteilung ist die Standardeinstellung. Die folgenden Optionen sind nur verfügbar, wenn entweder die Option **Bayes-Faktor schätzen** oder die Option **Beide Methoden verwenden** Bayes-Analyse ausgewählt ist.

### Poisson-Modell

Wenn diese Option ausgewählt ist, wird das Poisson-Modell für die beobachteten Daten angenommen.

### Multinomiales Modell

Wenn diese Option ausgewählt ist, wird das multinomiale Modell für die beobachteten Daten angenommen. Hierbei handelt es sich um die Standardeinstellung.

### Feste Ränder

Wählen Sie **Gesamtergebnis**, **Zeilensumme** oder **Spaltensumme** aus, um die festen Randsummen für die Kontingenztabelle anzugeben. **Gesamtergebnis** ist die Standardeinstellung.

### A-priori-Verteilung

Geben Sie den A-priori-Verteilungstyp bei der Schätzung des Bayes-Faktors an.

### Konjugieren

Wählen Sie diese Option aus, um eine konjugierte A-priori-Verteilung anzugeben. Verwenden Sie die Tabelle **Formparameter**, um die Formparameter  $a_{rs}$  für die Gammaverteilung anzugeben. Sie müssen die Formparameter angeben, wenn **Konjugieren** als A-priori-Verteilungstyp ausgewählt ist.

Wenn ein einzelner Wert angegeben wird, wird davon ausgegangen, dass alle  $a_{rs}$ -Werte gleich diesem Wert sind.  $a_{rs} = 1$  ist die Standardeinstellung. Wenn Sie mehr als einen Wert angeben müssen, können Sie die Werte durch Leerzeichen trennen.

Die Anzahl der numerischen Werte, die in den einzelnen Zeilen und Spalten angegeben sind, muss mit der Dimension der Kontingenztabelle übereinstimmen. Alle angegebenen Werte müssen  $> 0$  sein.

Klicken Sie auf **Zurücksetzen**, um die Werte zu löschen.

### Skalenparameter

Geben Sie den Skalenparameter  $b$  für Gammaverteilung an. Sie müssen einen einzelnen Wert  $> 0$  angeben.

### Mischungs-Dirichlet

Wählen Sie diese Option aus, um eine Mischungs-Dirichlet-A-priori-Verteilung anzugeben.

### Intrinsisch

Wählen Sie diese Option aus, um eine intrinsische A-priori-Verteilung anzugeben.

## Nicht parametrisches Modell

Wenn diese Option ausgewählt ist, wird das nicht parametrische Modell für die beobachteten Daten angenommen.

### Feste Ränder

Wählen Sie **Zeilensumme** oder **Spaltensumme** aus, um die festen Randsummen für die Kontingenztafel anzugeben. **Zeilensumme** ist die Standardeinstellung.

### A-priori-Verteilung

Geben Sie die Parameter für die Dirichlet-A-priori-Wahrscheinlichkeiten an. Sie müssen die Parameter für **A-priori-Verteilung** angeben, wenn **Nicht parametrisches Modell** ausgewählt ist. Wenn ein einzelner Wert angegeben wird, wird angenommen, dass alle  $\lambda_s$  gleich diesem Wert sind.  $\lambda_s = 1$  ist die Standardeinstellung. Wenn Sie mehr als einen Wert angeben müssen, können Sie die Werte durch Leerzeichen trennen. Alle angegebenen Werte müssen  $> 0$  sein. Die Anzahl der angegebenen, numerischen Werte muss mit der Dimension der Zeile oder Spalte übereinstimmen, die für die Kontingenztafel nicht als "Fest" definiert ist.

Klicken Sie auf **Zurücksetzen**, um die Werte zu löschen.

## Bayessche loglineare Regressionsmodelle: Drucken

Sie können angeben, wie der Inhalt in den Ausgabetafeln angezeigt werden soll.

### Tabellendesign

#### Tabelle unterdrücken

Wenn diese Option ausgewählt ist, wird die Kontingenztafel nicht in die Ausgabe eingeschlossen. Diese Einstellung ist standardmäßig nicht aktiviert.

**Anmerkung:** Die folgenden Einstellungen haben keine Auswirkung, wenn die Einstellung **Tabelle unterdrücken** aktiviert ist.

### Statistik

Geben Sie die Statistiken für Tests auf Unabhängigkeit an.

#### Chi-Quadrat

Wählen Sie diese Option aus, um die Chi-Quadrat-Statistik nach Pearson, Freiheitsgrade und die zweiseitige asymptotische Signifikanz zu berechnen. Für eine 2x2-Kontingenztafel berechnet diese Einstellung auch die Kontinuitätskorrekturstatistik nach Yates, Freiheitsgrade und die zugehörige zweiseitige asymptotische Signifikanz. Für eine 2x2-Kontingenztafel mit mindestens einer erwarteten Zellenanzahl  $< 5$  berechnet diese Einstellung auch die zweiseitige und einseitige exakte Signifikanz des exakten Tests nach Fisher.

#### Likelihood-Quotient

Wählen Sie diese Option aus, um die Teststatistik des Likelihood-Quotienten, Freiheitsgrade und die zugehörige zweiseitige asymptotische Signifikanz zu berechnen.

### Option 'counts'

Geben Sie an, welche Häufigkeitstypen in der Kontingenztafel enthalten sind.

#### Beobachtet

Wählen Sie diese Option aus, um die beobachtete Anzahl der Zellen in die Kontingenztafel aufzunehmen.

#### Erwartet

Wählen Sie diese Option aus, um die erwartete Anzahl der Zellen in die Kontingenztafel aufzunehmen.

### Prozentsätze

Geben Sie an, welche Prozentsatztypen in der Kontingenztafel enthalten sind.

#### Zeile

Wählen Sie diese Option aus, um Zeilenprozentsätze in die Kontingenztafel aufzunehmen.

#### Spalte

Wählen Sie diese Option aus, um Spaltenprozentsätze in die Kontingenztafel aufzunehmen.

## Summe

Wählen Sie diese Option aus, um Gesamtprozentsätze in die Kontingenztabelle aufzunehmen.

## Bayessche einfaktorielle ANOVA mit Messwiederholung: Modelle

Für diese Funktion sind SPSS Statistics Standard Edition oder die Option "Advanced Statistics" erforderlich.

In Modellen für bayessche einfaktorielle ANOVA mit Messwiederholung wird vorausgesetzt, dass es einen Messwert pro Subjekt gibt. Diese Voraussetzung wird allerdings nicht immer erfüllt. Es ist nicht ungewöhnlich, dass ein Studienkonzept die Untersuchung der Mittelwerte der abhängigen Variablen zu mehreren Zeitpunkten oder unter mehreren Bedingungen beinhaltet. Die bayessche einfaktorielle ANOVA mit Messwiederholung misst einen Faktor desselben Subjekts zu bestimmten Zeitpunkten oder unter bestimmten Bedingungen und Subjekte dürfen sich innerhalb der Messniveaus überkreuzen. Es wird angenommen, dass jedes Subjekt eine einzelne Beobachtung zu jedem Zeitpunkt und jeder Bedingung hat (die Interaktion bei der Behandlung des Subjekts wird nicht berücksichtigt).

1. Wählen Sie Folgendes in den Menüs aus:

**Analysieren > Bayes-Statistik > Einfaktorielle ANOVA mit Messwiederholung**

2. Wählen Sie in der Liste **Verfügbare Variablen** mindestens zwei Messwiederholungsvariablen aus.

3. Wählen Sie optional aus der Liste **Verfügbare Variablen** eine einzelne Variable aus, die als Regression **Gewichtung** dienen soll. Das Variablenfeld **Gewichtung** kann leer sein.

**Anmerkung:** Die Liste mit verfügbaren Variablen enthält alle Variablen außer Zeichenfolgevariablen.

4. Wählen Sie die gewünschte **Bayes-Analyse** aus:

- **A-posteriori-Verteilung charakterisieren:** Bei Auswahl dieser Option wird die Bayessche Inferenz aus einer Perspektive erstellt, die durch die Charakterisierung von A-posteriori-Verteilungen erreicht wird. Sie können die marginale A-posteriori-Verteilung der relevanten Parameter untersuchen, indem Sie die übrigen störenden Parameter herausintegrieren und außerdem Zuverlässigkeitsintervalle erstellen, um die direkte Inferenz zu zeichnen. Dies ist die Standardeinstellung.
- **Bayes-Faktor schätzen:** Bei Auswahl dieser Option wird durch Schätzen von Bayes-Faktoren (eine der bemerkenswerten Methodiken der bayesschen Inferenz) ein natürlicher Quotient zum Vergleichen der marginalen Wahrscheinlichkeiten zwischen einer Nullhypothese und einer Alternativhypothese gebildet.

*Tabelle 11. Häufig verwendete Schwellenwerte zum Definieren der Signifikanz von Nachweisen*

Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie	Bayes-Faktor	Nachweiskategorie
>100	Extremer Nachweis für H1	1-3	Anekdotische Evidenz für H1	1/30-1/10	Starker Nachweis für H0
30-100	Sehr starker Nachweis für H1	1	Keine Evidenz	1/100-1/30	Sehr starker Nachweis für H0
10-30	Starke Evidenz für H1	1/3-1	Anekdotische Evidenz für H0	1/100	Extremer Nachweis für H0
3-10	Moderate Evidenz für H1	1/10-1/3	Moderate Evidenz für H0		

*H0: Nullhypothese*

*H1: Alternativhypothese*

- **Beide Methoden verwenden:** Bei Auswahl dieser Option werden beide Inferenzmethoden **A-posteriori-Verteilung charakterisieren** und **Bayes-Faktor schätzen** verwendet.

Die folgenden Optionen sind verfügbar:

- Klicken Sie auf **Kriterien**, um die Einstellungen für das Zuverlässigkeitsintervall in Prozent und die numerische Methode anzugeben.
- Klicken Sie auf **Bayes-Faktor**, um Einstellungen für Bayes-Faktoren anzugeben.
- Klicken Sie auf **Diagramme**, um die A-posteriori-Verteilungen des Gruppenmittelwerts dazustellen.

## Bayessche einfaktorielle ANOVA mit Messwiederholung: Kriterien

Sie können die folgenden Analyse Kriterien für bayessche einfaktorielle ANOVA-Modelle mit Messwiederholung angeben:

### Zuverlässigkeitsintervall in %

Geben Sie das Signifikanzniveau für die Berechnung von Zuverlässigkeitsintervallen an. Das Standardniveau ist 95 %.

### Numerische Methode

Geben Sie die numerische Methode an, die zum Schätzen des Integrals verwendet wird.

### Benutzerdefinierten Startwert festlegen

Bei Auswahl dieser Option können Sie im Feld **Start** einen benutzerdefinierten Startwert angeben. Der Standardwert ist 2.000.000. Der Wert muss eine positive Ganzzahl zwischen 1 und 2.147.483.647 sein. Standardmäßig wird ein Startwert für Zufallszahlen zugewiesen.

### Anzahl der Monte-Carlo-Stichproben

Geben Sie die Anzahl der Punkte an, für die Stichproben für die Monte-Carlo-Approximation gezogen werden sollen. Der Wert muss eine positive Ganzzahl zwischen  $10^3$  und  $10^6$  sein. Der Standardwert ist 30.000.

## Bayessche einfaktorielle ANOVA mit Messwiederholung: Bayes-Faktor

Sie können die Methode zum Schätzen des Bayes-Faktors für Modelle mit bayesscher einfaktorielle ANOVA mit Messwiederholung angeben. Die folgenden Optionen sind nur verfügbar, wenn entweder die Option **Bayes-Faktor schätzen** oder die Option **Beide Methoden verwenden** Bayes-Analyse ausgewählt ist.

### Bayes-Informationskriterien (BIC)

Verwendet eine Erweiterung der BIC-Approximation für das Design der Messwiederholungen, um Bayes-Faktoren zu schätzen. Die Einstellung leitet den effektiven Stichprobenumfang ab, der für die Korrelation mit Messwiederholungen verwendet wird, und schlägt einen verbesserten Penalisierungsterm vor, wenn die BIC für die Auswahl zwischen zwei konkurrierenden Modellen geschätzt werden. Hierbei handelt es sich um die Standardeinstellung.

### Rouder's gemischtes Design

Verwendet die multivariaten Verallgemeinerungen der Cauchy-Verteilung als A-Priori-Verteilung für die standardisierte Effektgröße und eine nicht informative A-priori-Verteilung für die Varianz.

**Anmerkung:** Die Einstellung für die globale Häufigkeitsgewichtung und die Regressionsgewichtung werden ignoriert, wenn diese Option ausgewählt wird.

## Bayessche einfaktorielle ANOVA mit Messwiederholung: Diagramme

Sie können die ausgegebenen Diagramme steuern, um die A-posteriori-Verteilungen der Gruppenmittelwerte zu veranschaulichen. In der Tabelle werden alle Variablen aufgelistet, die im Dialogfeld "Variablen"

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

<sup>20</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

als Messwiederholungen ausgewählt sind. Wählen Sie die darzustellenden Messwiederholungsvariablen aus.

## Kernel-Ridge-Regression

---

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

### Beispiel

#### Statistik

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

### Erläuterung der Daten

#### Daten

- Sie können eine beliebige oder alle der acht verschiedenen Kernelfunktionen angeben.
- Die ausgewählte Kernelfunktion bestimmt, welche Hyperparameter aktiv sind.
- Zu den Hyperparametern gehören Alpha für die Ridge-Regularisierung, die allen Kernen gemeinsam sind, plus drei weitere Hyperparameter für jede spezifische Kernel-Funktion.
- Wenn mehrere Kernel-Unterbefehle angegeben werden oder mehr als ein Wert für einen beliebigen Parameter angegeben wird, wird eine Rastersuche mit der Quervalidierung für die Auswertung von Modellen durchgeführt und das beste passende Modell, das auf "ferngehaltenen" Daten basiert, ausgewählt.
- Die Erweiterung nimmt getrennte Variablen aus der Prozedur "Datei aufteilen" an und gewichtet diese anhand der Prozedur "Fälle gewichten".
- Wenn Gewichtungen eingeschlossen werden, werden sie bei der Erstellung von angepassten Werten in allen Analysen verwendet. Aufgrund von Einschränkungen in der Bewertungsmethode in der Klasse `sklearn.model_selection.GridSearchCV` werden Kreuzvalidierungsauswertungen, die für die Modellauswahl verwendet werden, nicht gewichtet.

#### Voraussetzungen

### Abrufen einer Kernel-Ridge-Regression

1. Wählen Sie Folgendes in den Menüs aus:

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

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

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

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

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

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

- a. Klicken Sie auf das Steuerelement "Hinzufügen" (+), um zusätzliche Kernelfunktionen aufzunehmen.
- b. Klicken Sie auf die leere Zelle in der Spalte **Kernel**, um eine Kernelfunktion auszuwählen.
- c. Klicken Sie doppelt auf eine beliebige Kernelfunktionszeilenzelle, um die Werte für die Kernelfunktionsparameter für die entsprechende Spalte (**Alpha**, **Gamma**, **Coef0**, **Grad**) anzugeben. Weitere Informationen finden Sie unter „Kernelparameter“ auf Seite 144. Die Standardparameter für die Optimierung der Kernelfunktion sind im Folgenden aufgelistet.

**Additive\_CHI2**

ALPHA=1 GAMMA=1

**CHI2**

ALPHA=1 GAMMA=1

**Kosinus**

ALPHA=1

**Laplace**

ALPHA=1 GAMMA=1/p

**Linear**

Die Standard-Kernelfunktion. ALPHA=1

**Polynomial**

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

**RBF**

ALPHA=1 GAMMA=1/p

**Sigma**

ALPHA=1 GAMMA=1/p COEF0=1

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

5. Klicken Sie optional auf **Optionen** , um die Anzahl der Kreuzvalidierungsfalten, Anzeigeoptionen, Ploteinstellungen und zu speichernde Elemente anzugeben. Weitere Informationen finden Sie unter „Kernel-Ridge-Regression: Optionen“ auf Seite 144.
6. Klicken Sie auf **OK**.

## Kernelparameter

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

### Einzelne Parameter angeben

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

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

### Rasterparameter angeben

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

## Kernel-Ridge-Regression: Optionen

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

### **Anzahl der Kreuzvalidierungsfalten**

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

### **Anzeige**

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

### **Empfohlen**

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

### **Vergleichen**

Zeigt grundlegende Ergebnisse für alle ausgewerteten Modelle an.

### **Modelle und Falten vergleichen**

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

### **Diagramm**

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

### **Beobachtet vs. vorhergesagt**

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

### **Residuen vs. Vorhergesagt**

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

### **Speichern**

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

### **Vorhergesagte Werte**

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

### **Residuen**

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

### **Duale Koeffizienten**

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



## Bemerkungen

---

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

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

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

*IBM Director of Licensing*

*IBM Europe, Middle East & Africa*

*North Castle Drive, MD-NC119*

*Armonk, NY 10504-1785*

*France*

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

*Intellectual Property Licensing*

*Legal and Intellectual Property Law*

*IBM Japan Ltd.*

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

*Tokyo 103-8510, Japan*

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

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

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

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

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

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

*IBM Director of Licensing*

*IBM Europe, Middle East & Africa*

*North Castle Drive, MD-NC119*

*Armonk, NY 10504-1785*

*France*

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

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

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

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

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

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

#### COPYRIGHTLIZENZ:

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

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

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

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

## Marken

---

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

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

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

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

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

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

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

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



# Index

## A

- A-priori-Zufallseffekte
  - in Varianzkomponenten [21](#)
- Abweichungsresiduen
  - in "Verallgemeinerte lineare Modelle" [39](#)
- Allgemeine loglineare Analyse
  - Anzeigeoptionen [72](#)
  - Faktoren [70](#)
  - Grafik [72](#)
  - Konfidenzintervalle [72](#)
  - Kontraste [70](#)
  - Kriterien [72](#)
  - Modellspezifikationen [72](#)
  - Residuen [73](#)
  - Speichern von Variablen [73](#)
  - Speichern von vorhergesagten Werten [73](#)
  - Verteilung der Zellenhäufigkeiten [70](#)
  - Zellenkovariaten [70](#)
  - Zellenstrukturen [70](#)
  - zusätzliche Funktionen beim Befehl [73](#)
- Allgemeine schätzbare Funktion
  - in "Verallgemeinerte Schätzungsgleichungen" [49](#)
- allgemeine schätzbare Funktion(en)
  - in "Verallgemeinerte lineare Modelle" [37](#)
- Analyse der Kovarianz
  - in "GLM - Multivariat" [1](#)
- angepasste Modelle
  - in GLM - Messwiederholungen [12](#)
- ANOVA
  - in "GLM - Multivariat" [1](#)
- Anpassungsgüte
  - in "Verallgemeinerte lineare Modelle" [37](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [49](#)
- Ausgeschlossene Residuen
  - in GLM [7](#)
  - in GLM - Messwiederholung [17](#)

## B

- Binominalverteilung
  - in "Verallgemeinerte lineare Modelle" [30](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [43](#)
- Bonferroni
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Breslow-Test
  - in Kaplan-Meier [97](#)

## C

- Cook-Distanz
  - in "Verallgemeinerte lineare Modelle" [39](#)
  - in GLM [7](#)
  - in GLM - Messwiederholung [17](#)
- Cox-Regression
  - Basisfunktionen [101](#)

- Cox-Regression (*Forts.*)
  - Beispiel [99](#)
  - DfBeta [101](#)
  - Ereignis definieren [102](#)
  - Grafik [100](#)
  - Hazardfunktion [101](#)
  - Iterationen [101](#)
  - kategoriale Kovariaten [99](#)
  - Kontraste [99](#)
  - Kovariaten [99](#)
  - partielle Residuen [101](#)
  - schrittweise Aufnahme und Ausschluss [101](#)
  - Speichern von neuen Variablen [101](#)
  - Statistik [99](#), [101](#)
  - Statusvariable für Überlebensverteilung [102](#)
  - Überlebensfunktion [101](#)
  - Zeichenfolge-Kovariaten [99](#)
  - zeitabhängige Kovariaten [102](#)
  - zusätzliche Funktionen beim Befehl [102](#)

## D

- Deskriptive Statistiken
  - in "Lineare gemischte Modelle" [28](#)
  - in "Verallgemeinerte lineare Modelle" [37](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [49](#)
- Dunnett-C
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Dunnett-T3
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Dunnett-Test
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)

## E

- Eingeschränkte Maximum-Likelihood-Schätzung
  - in Varianzkomponenten [21](#)
- Erstellen von Termen [3](#), [13](#), [21](#), [70](#), [72](#), [75](#)

## F

- Faktoren
  - in GLM - Messwiederholung [12](#)
- Faktorstufeninformationen
  - in "Lineare gemischte Modelle" [28](#)
- Feste Effekte
  - in "Lineare gemischte Modelle" [25](#)
- Feste vorhergesagte Werte
  - in "Lineare gemischte Modelle" [29](#)
- Fisher-Scoring
  - in "Lineare gemischte Modelle" [27](#)

## G

- Gammaverteilung
  - in "Verallgemeinerte lineare Modelle" [30](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [43](#)
- Gehan-Test
  - in Sterbetafeln [78](#)
- Gemischte Modelle
  - linear [22](#)
  - verallgemeinerte lineare gemischte Modelle [53](#)
- Geringste signifikante Differenz
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Gesättigte Modelle
  - in "Modellauswahl für loglineare Analyse" [69](#)
  - in GLM - Messwiederholungen [12](#)
  - in Varianzkomponenten [20](#)
- geschätzte Randmittel
  - in "Lineare gemischte Modelle" [29](#)
  - in "Verallgemeinerte lineare Modelle" [38](#)
- Geschätzte Randmittel
  - in "Verallgemeinerte Schätzungsgleichungen" [50](#)
- Gewichtete Schätzwerte
  - in GLM [7](#)
  - in GLM - Messwiederholung [17](#)
- GLM
  - Speichern von Matrizen [7](#)
  - Speichern von Variablen [7](#)
- GLM - Messwiederholung
  - Profilplots [14](#)
- GLM - Messwiederholungen
  - Faktoren definieren [12](#)
  - Modell [12](#)
  - Post-hoc-Tests [15](#)
  - Speichern von Variablen [17](#)
  - zusätzliche Funktionen beim Befehl [18](#)
- GLM - Multivariat
  - abhängige Variable [1](#)
  - Faktoren [1](#)
  - Kovariaten [1](#)
  - Post-hoc-Tests [5](#)
  - Profilplots [5](#)
- GLOR
  - in "Allgemeine loglineare Analyse" [70](#)
- Grafik
  - in "Allgemeine loglineare Analyse" [72](#)
  - in "Logit-loglineare Analyse" [75](#)
- GT2 nach Hochberg
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)

## H

- Häufigkeiten
  - in "Modellauswahl für loglineare Analyse" [70](#)
- Hazardrate
  - in Sterbetafeln [76](#)
- Hebelwerte
  - in "Verallgemeinerte lineare Modelle" [39](#)
  - in GLM [7](#)
  - in GLM - Messwiederholung [17](#)
- Hierarchische loglineare Modelle [68](#)
- Hierarchische Modelle
  - verallgemeinerte lineare gemischte Modelle [53](#)

- Hierarchische Zerlegung
  - in Varianzkomponenten [21](#)

## I

- Identitätsverknüpfungsfunktion
  - in "Verallgemeinerte lineare Modelle" [30](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [43](#)
- Interaktionsterme
  - in "Lineare gemischte Modelle" [25](#)
- Inverse Normalverteilung
  - in "Verallgemeinerte lineare Modelle" [30](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [43](#)
- Iteration
  - in "Modellauswahl für loglineare Analyse" [70](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [47](#)
- Iterationen
  - in "Verallgemeinerte lineare Modelle" [35](#)
- Iterationsverlauf
  - in "Lineare gemischte Modelle" [27](#)
  - in "Verallgemeinerte lineare Modelle" [37](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [49](#)

## K

- Kaplan-Meier
  - Beispiel [96](#)
  - Definieren von Ereignissen [97](#)
  - Grafik [98](#)
  - linearer Trend für Faktorstufen [97](#)
  - Mittelwert und Median der Überlebenszeit [98](#)
  - Quartile [98](#)
  - Speichern von neuen Variablen [98](#)
  - Statistik [96, 98](#)
  - Statusvariablen für Überlebensverteilung [97](#)
  - Überlebenstabellen [98](#)
  - Vergleichen von Faktorstufen [97](#)
  - zusätzliche Funktionen beim Befehl [98](#)
- Kernel Ridge
  - alpha [143](#)
  - coef0 [143](#)
  - Einzelmodell [143](#)
  - Gamma [143](#)
  - Grad [143](#)
  - Modellauswahl [143](#)
- Kernel-Ridge-Regression
  - Anzeige [144](#)
  - Grafik [144](#)
  - Kreuzvalidierungsfalten [144](#)
  - Parameter [144](#)
  - Rasterparameter [144](#)
  - speichern [144](#)
- Konfidenzintervalle
  - in "Allgemeine loglineare Analyse" [72](#)
  - in "Lineare gemischte Modelle" [28](#)
  - in "Logit-loglineare Analyse" [75](#)
- Kontingenztabellen
  - in "Allgemeine loglineare Analyse" [70](#)
- Kontraste
  - in "Allgemeine loglineare Analyse" [70](#)
  - in "Logit-loglineare Analyse" [73](#)
  - in Cox-Regression [99](#)
- Kontrastkoeffizientenmatrix

Kontrastkoeffizientenmatrix (*Forts.*)  
 in "Verallgemeinerte lineare Modelle" [37](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [49](#)

Konvergenz der Hesse-Matrix  
 in "Verallgemeinerte lineare Modelle" [35](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [47](#)

Korrelationsmatrix  
 in "Lineare gemischte Modelle" [28](#)  
 in "Verallgemeinerte lineare Modelle" [37](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [49](#)

Kovarianzmatrix  
 in "Lineare gemischte Modelle" [28](#)  
 in "Verallgemeinerte lineare Modelle" [35, 37](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [47, 49](#)  
 in GLM [7](#)

Kovarianzmatrix der Zufallseffekte  
 in "Lineare gemischte Modelle" [28](#)

Kovarianzparametertest  
 in "Lineare gemischte Modelle" [28](#)

Kovarianzstrukturen  
 in "Lineare gemischte Modelle" [107](#)

Kovariaten  
 in Cox-Regression [99](#)

Kreuztabulierung  
 in "Modellauswahl für loglineare Analyse" [68](#)

Kumulative Cauchit-Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Kumulative Logit-Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Kumulative Probit-Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

kundenspezifische Modelle  
 in "Modellauswahl für loglineare Analyse" [69](#)  
 in Varianzkomponenten [20](#)

## L

L-Matrix  
 in "Verallgemeinerte lineare Modelle" [37](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [49](#)

Lagrange-Multiplikator-Test  
 in "Verallgemeinerte lineare Modelle" [37](#)

Likelihood-Residuen  
 in "Verallgemeinerte lineare Modelle" [39](#)

Lineare gemischte Modelle  
 Erstellen von Termen [25, 26](#)  
 Feste Effekte [25](#)  
 geschätzte Randmittel [29](#)  
 Interaktionsterme [25](#)  
 Kovarianzstruktur [107](#)  
 Modell [28](#)  
 Schätzkriterien [27](#)  
 Variablen speichern [29](#)  
 zufällige Effekte [26, 30](#)  
 zusätzliche Funktionen beim Befehl [30](#)

Log-Likelihood-Konvergenz.  
 in "Lineare gemischte Modelle" [27](#)  
 in "Verallgemeinerte lineare Modelle" [35](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [47](#)

Log-Log komplementär, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)

Log-Log komplementär, Verknüpfungsfunktion (*Forts.*)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Log-Log kumulativ komplementär, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Log-Log kumulativ negativ, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Log-Log negativ, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Log-Rang-Test  
 in Kaplan-Meier [97](#)

Log-Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Log. Komplement, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

logistische Regression  
 verallgemeinerte lineare gemischte Modelle [53](#)

Logit-loglineare Analyse  
 Anzeigoptionen [75](#)  
 Faktoren [73](#)  
 Grafik [75](#)  
 Konfidenzintervalle [75](#)  
 Kontraste [73](#)  
 Kriterien [75](#)  
 Modellspezifikation [74](#)  
 Residuen [76](#)  
 Speichern von Variablen [76](#)  
 Verteilung der Zellenhäufigkeiten [73](#)  
 Vorhergesagte Werte [76](#)  
 Zellenkovariaten [73](#)  
 Zellenstrukturen [73](#)

Logit-Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)

Loglineare Analyse  
 Allgemeine loglineare Analyse [70](#)  
 in "Verallgemeinerte lineare gemischte Modelle" [53](#)  
 Logit-loglineare Analyse [73](#)

Longitudinalmodelle  
 verallgemeinerte lineare gemischte Modelle [53](#)

LSD nach Fisher  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [15](#)

## M

Maximum-Likelihood-Schätzung  
 in Varianzkomponenten [21](#)

Mehrebenenmodelle  
 verallgemeinerte lineare gemischte Modelle [53](#)

MINQUE  
 in Varianzkomponenten [21](#)

Modell für loglineare Analyse auswählen  
 Definieren der Faktorenbereiche [69](#)  
 Optionen [70](#)

Modellansicht  
 in "Verallgemeinerte lineare gemischte Modelle" [64](#)

Modellauswahl für loglineare Analyse  
 -Modelle [69](#)

Modellauswahl für loglineare Analyse (*Forts.*)  
 zusätzliche Funktionen beim Befehl [70](#)  
 Modellbildende Klasse  
 in "Modellauswahl für loglineare Analyse" [69](#)  
 Modellinformationen  
 in "Verallgemeinerte lineare Modelle" [37](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [49](#)  
 Multinomiale logistische Regression  
 verallgemeinerte lineare gemischte Modelle [53](#)  
 Multinomiale Logit-Modelle [73](#)  
 Multinomialverteilung  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Multipler Spannweitentest nach Duncan  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [15](#)  
 Multipler Spannweitentest nach Ryan-Einot-Gabriel-Welsch  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [15](#)  
 Multivariat: ANOVA [1](#)  
 Multivariat: GLM [1](#)  
 Multivariate Regression [1](#)

## N

Negativ binomial, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Negative Binomialverteilung  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Newman-Keuls  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [15](#)  
 Newton-Raphson-Methode  
 in "Allgemeine loglineare Analyse" [70](#)  
 in "Logit-loglineare Analyse" [73](#)  
 Nicht standardisierte Residuen  
 in GLM [7](#)  
 in GLM - Messwiederholung [17](#)  
 Normalverteilung  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Normalverteilungsdiagramme  
 in "Modellauswahl für loglineare Analyse" [70](#)

## O

Odds Power, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Odds-Verhältnis  
 in "Allgemeine loglineare Analyse" [70](#)

## P

Paarweiser Vergleichstest nach Gabriel  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [15](#)  
 Paarweiser Vergleichstest nach Games und Howell  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [15](#)  
 Parameterkonvergenz

Parameterkonvergenz (*Forts.*)  
 in "Lineare gemischte Modelle" [27](#)  
 in "Verallgemeinerte lineare Modelle" [35](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [47](#)  
 Parameterkovarianzmatrix  
 in "Lineare gemischte Modelle" [28](#)  
 Parameterschätzungen  
 in "Allgemeine loglineare Analyse" [70](#)  
 in "Lineare gemischte Modelle" [28](#)  
 in "Logit-loglineare Analyse" [73](#)  
 in "Modellauswahl für loglineare Analyse" [70](#)  
 in "Verallgemeinerte lineare Modelle" [37](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [49](#)  
 Parametrische beschleunigte Fehlerzeitmodelle  
 drucken [82](#)  
 Export [84](#)  
 Kriterien [79](#)  
 modell [80](#)  
 plotten [83](#)  
 Schätzung [81](#)  
 und Einrichtungen [78](#)  
 vorhersagen [83](#)  
 Parametrische Frailty-Modelle  
 Statusvariablen für Überlebensverteilung [91](#)  
 Parametrische gemeinsam genutzte Frailty-Modelle  
 drucken [89](#)  
 Export [91](#)  
 Kriterien [86](#)  
 modell [87](#)  
 plotten [91](#)  
 Schätzung [88](#)  
 und Einrichtungen [85](#)  
 vorhersagen [90](#)  
 Pearson-Residuen  
 in "Verallgemeinerte lineare Modelle" [39](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [51](#)  
 Poisson-Regression  
 in "Allgemeine loglineare Analyse" [70](#)  
 verallgemeinerte lineare gemischte Modelle [53](#)  
 Poisson-Verteilung  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Power, Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Probit-Analyse  
 verallgemeinerte lineare gemischte Modelle [53](#)  
 Probit-Verknüpfungsfunktion  
 in "Verallgemeinerte lineare Modelle" [30](#)  
 in "Verallgemeinerte Schätzungsgleichungen" [43](#)  
 Profilplots  
 in "GLM - Multivariat" [5](#)  
 in GLM - Messwiederholung [14](#)  
 Proportional-Hazards-Modell  
 in Cox-Regression [99](#)

## Q

Quadratsumme [3](#), [13](#)  
 Quadratsummen  
 in "Lineare gemischte Modelle" [26](#)  
 in Varianzkomponenten [21](#)

## R

- R-E-G-W F
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- R-E-G-W Q
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Referenzkategorie
  - in "Verallgemeinerte lineare Modelle" [33](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [46](#)
- Residuen
  - in "Allgemeine loglineare Analyse" [73](#)
  - in "Lineare gemischte Modelle" [29](#)
  - in "Logit-loglineare Analyse" [76](#)
  - in "Modellauswahl für loglineare Analyse" [70](#)
  - in "Verallgemeinerte lineare Modelle" [39](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [51](#)
- Residuenkovarianzmatrix
  - in "Lineare gemischte Modelle" [28](#)
- Rückwärtselimination
  - in "Modellauswahl für loglineare Analyse" [68](#)
- Ryan-Einot-Gabriel-Welsch multiple F
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)

## S

- Scheffé-Test
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Schritthalbierungen
  - in "Lineare gemischte Modelle" [27](#)
  - in "Verallgemeinerte lineare Modelle" [35](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [47](#)
- Scoring
  - in "Lineare gemischte Modelle" [27](#)
- Segmentierte zeitabhängige Kovariaten
  - in Cox-Regression [102](#)
- Sidak-Test
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Skalenparameter
  - in "Verallgemeinerte lineare Modelle" [35](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [47](#)
- Standardfehler
  - in GLM [7](#)
  - in GLM - Messwiederholung [17](#)
- Standardisierte Residuen
  - in GLM [7](#)
  - in GLM - Messwiederholung [17](#)
- Sterbetafeln
  - Anzeige von Tabellen unterdrücken [78](#)
  - Beispiel [76](#)
  - Faktorvariablen [78](#)
  - Grafik [78](#)
  - Hazardrate [76](#)
  - Statistik [76](#)
  - Statusvariablen für Überlebensverteilung [84](#)
  - Überlebensfunktion [76](#)
  - Vergleichen von Faktorstufen [78](#)
  - Wilcoxon-Test (Gehan-Test) [78](#)
  - zusätzliche Funktionen beim Befehl [78](#)
- Student-Newman-Keuls-Prozedur

- Student-Newman-Keuls-Prozedur (Forts.)
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Subjektvariablen
  - in "Lineare gemischte Modelle" [24](#)

## T

- Tamhane-T2
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Tarone-Ware-Test
  - in Kaplan-Meier [97](#)
- Toleranz für Prüfung auf Singularität
  - in "Lineare gemischte Modelle" [27](#)
- Trennung
  - in "Verallgemeinerte lineare Modelle" [35](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [47](#)
- Tukey-B-Test
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Tukey-HSD-Test
  - in "GLM - Multivariat" [5](#)
  - in GLM - Messwiederholung [15](#)
- Tweedie-Verteilung
  - in "Verallgemeinerte lineare Modelle" [30](#)
  - in "Verallgemeinerte Schätzungsgleichungen" [43](#)

## U

- Überleben AFT
  - Überlebensdialog-Kategorievariablen [84](#)
- Überlebensanalyse
  - in Cox-Regression [99](#)
  - in der Kernel Ridge Regression [143](#)
  - in Kaplan-Meier [96](#)
  - in Sterbetafeln [76](#)
  - zeitabhängige Cox-Regression [102](#)
- Überlebensfunktion
  - in Sterbetafeln [76](#)

## V

- VA
  - in GLM - Messwiederholung [10](#)
- Variablen für Messwiederholungen
  - in "Lineare gemischte Modelle" [24](#)
- Varianzanalyse
  - in "Verallgemeinerte lineare gemischte Modelle" [53](#)
  - in Varianzkomponenten [21](#)
- Varianzkomponenten
  - Modell [20](#)
  - Optionen [21](#)
  - Speichern von Ergebnissen [22](#)
  - zusätzliche Funktionen beim Befehl [22](#)
- verallgemeinerte lineare gemischte Modelle
  - Datenstruktur [65](#)
  - feste Koeffizienten [66](#)
  - Klassifikationstabelle [65](#)
  - Kovarianzparameter [67](#)
  - Modellansicht [64](#)
  - Modellexport [63](#), [64](#)
  - Speichern von Feldern [63](#), [64](#)

## Verallgemeinerte lineare gemischte Modelle

- Analysegewichtung [61](#)
- benutzerdefinierte Terme [59](#)
- Block für zufällige Effekte [59](#)
- Feste Effekte [58](#), [65](#)
- geschätzte Mittelwerte [67](#)
- geschätzte Randmittel [62](#)
- Kovarianzen der Zufallseffekte [66](#)
- Modellzusammenfassung [64](#)
- Offset [61](#)
- Verknüpfungsfunktion [56](#)
- Vorhersage nach Beobachtung [65](#)
- Zielverteilung [56](#)
- zufällige Effekte [59](#)

## Verallgemeinerte lineare Modelle

- Anfangswerte [36](#)
- Antwort [33](#)
- geschätzte Randmittel [38](#)
- in "Verallgemeinerte lineare gemischte Modelle" [53](#)
- Modellexport [40](#)
- Modellspezifikation [34](#)
- Modelltypen [30](#)
- Optionen für kategoriale Faktoren [34](#)
- Prädiktoren [34](#)
- Referenzkategorie für binäre Antwort [33](#)
- Schätzkriterien [35](#)
- Speichern von Variablen im aktiven Dataset [39](#)
- Statistikdaten [37](#)
- Verknüpfungsfunktion [30](#)
- Verteilung [30](#)

## Verallgemeinerte Schätzungsgleichungen

- Anfangswerte [49](#)
- Antwort [45](#)
- Geschätzte Randmittel [50](#)
- Modellexport [52](#)
- Modellspezifikationen [47](#)
- Optionen für kategoriale Faktoren [46](#)
- Prädiktoren [46](#)
- Referenzkategorie für binäre Antwort [46](#)
- Schätzkriterien [47](#)
- Speichern von Variablen im aktiven Dataset [51](#)
- Statistik [49](#)
- Typ des Modells [43](#)

## Verallgemeinertes lineares Modell

- verallgemeinerte lineare gemischte Modelle [53](#)

## Verallgemeinertes Log-Odds-Verhältnis

- in "Allgemeine loglineare Analyse" [70](#)

## Verknüpfungsfunktion

- Verallgemeinerte lineare gemischte Modelle [56](#)

## Verschachtelte Terme

- in "Lineare gemischte Modelle" [26](#)
- in "Verallgemeinerte lineare Modelle" [34](#)
- in "Verallgemeinerte Schätzungsgleichungen" [47](#)

## vorhergesagte Werte

- in "Lineare gemischte Modelle" [29](#)

## Vorhergesagte Werte

- in "Allgemeine loglineare Analyse" [73](#)
- in "Logit-loglineare Analyse" [76](#)

## W

### Wald-Statistik

- in "Allgemeine loglineare Analyse" [70](#)
- in "Logit-loglineare Analyse" [73](#)

## Waller-Duncan-Test

- in "GLM - Multivariat" [5](#)
- in GLM - Messwiederholung [15](#)

## Wilcoxon-Test

- in Sterbetafeln [78](#)

## Z

### Zeichenfolge-Kovariaten

- in Cox-Regression [99](#)

### Zensierte Fälle

- in Cox-Regression [99](#)
- in Kaplan-Meier [96](#)
- in Sterbetafeln [76](#)

### zufällige Effekte

- in "Lineare gemischte Modelle" [26](#), [30](#)

### Zusammenfassung der Fallverarbeitung

- in "Verallgemeinerte lineare Modelle" [37](#)
- in "Verallgemeinerte Schätzungsgleichungen" [49](#)



