

IBM SPSS Advanced Statistics 19



Note: Before using this information and the product it supports, read the general information under Notices auf S. 173.

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright SPSS Inc. 1989, 2010.**

IBM® SPSS® Statistics ist ein umfassendes System zum Analysieren von Daten. Das optionale Zusatzmodul Advanced Statistics bietet die zusätzlichen Analyseverfahren, die in diesem Handbuch beschrieben sind. Die Prozeduren im Zusatzmodul Advanced Statistics müssen zusammen mit SPSS Statistics Core verwendet werden. Sie sind vollständig in dieses System integriert.

Über SPSS Inc., ein Unternehmen von IBM

SPSS Inc., ein Unternehmen von IBM, ist ein führender globaler Anbieter von Analysesoftware und -lösungen zur Prognoseerstellung. Mit der vollständigen Produktpalette des Unternehmens – Datenerfassung, Statistik, Modellierung und Bereitstellung – werden Einstellungen und Meinungen von Personen erfasst und Ergebnisse von künftigen Interaktionen mit Kunden prognostiziert. Anschließend werden diese Erkenntnisse durch die Einbettung der Analysen in Geschäftsprozesse praktisch umgesetzt. Lösungen von SPSS Inc. sind durch die Konzentration auf die Zusammenführung von Analysefunktionen, IT-Architektur und Geschäftsprozessen für zusammenhängende unternehmensübergreifende Geschäftsziele konzipiert. Kunden aus den Bereichen Wirtschaft, Regierung und Wissenschaft vertrauen weltweit auf die Technologie von SPSS Inc. als Wettbewerbsvorteil, wenn es gilt, Kunden anzuziehen, zu binden und neue Kunden zu gewinnen und dabei Betrugsfälle zu verringern und Risiken zu entschärfen. SPSS Inc. wurde im Oktober 2009 von IBM übernommen. Weitere Informationen erhalten Sie unter <http://www.spss.com>.

Technischer Support

Kunden mit Wartungsvertrag können den Technischen Support in Anspruch nehmen. Kunden können sich an den Technischen Support wenden, wenn sie Hilfe bei der Arbeit mit den Produkten von SPSS Inc. oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Wie Sie den Technischen Support kontaktieren können, entnehmen Sie der Website von SPSS Inc. unter <http://support.spss.com>. Über die Website unter <http://support.spss.com/default.asp?refpage=contactus.asp> können Sie auch nach Ihrem örtlichen Büro suchen. Wenn Sie Hilfe anfordern, halten Sie bitte Informationen bereit, um sich, Ihre Organisation und Ihren Supportvertrag zu identifizieren.

Kundendienst

Wenden Sie sich bei Fragen zur Lieferung oder Ihrem Kundenkonto an Ihr regionales Büro, das Sie auf der Website unter <http://www.spss.com/worldwide> finden. Halten Sie bitte stets Ihre Seriennummer bereit.

Ausbildungsseminare

SPSS Inc. bietet öffentliche und unternehmensinterne Seminare an. Alle Seminare beinhalten auch praktische Übungen. Seminare finden in größeren Städten regelmäßig statt. Wenn Sie weitere Informationen zu diesen Seminaren wünschen, wenden Sie sich an Ihr regionales Büro, das Sie auf der Website unter <http://www.spss.com/worldwide> finden.

Weitere Veröffentlichungen

Die Handbücher *SPSS Statistics: Guide to Data Analysis*, *SPSS Statistics: Statistical Procedures Companion* und *SPSS Statistics: Advanced Statistical Procedures Companion*, die von Marija Norušis geschrieben und von Prentice Hall veröffentlicht wurden, werden als Quelle für Zusatzinformationen empfohlen. Diese Veröffentlichungen enthalten statistische Verfahren in den Modulen “Statistics Base”, “Advanced Statistics” und “Regression” von SPSS. Diese Bücher werden Sie dabei unterstützen, die Funktionen und Möglichkeiten von IBM® SPSS® Statistics optimal zu nutzen. Dabei ist es unerheblich, ob Sie ein Neuling im Bereich der Datenanalyse sind oder bereits über umfangreiche Vorkenntnisse verfügen und damit in der Lage sind, auch die erweiterten Anwendungen zu nutzen. Weitere Informationen zu den Inhalten der Veröffentlichungen sowie Auszüge aus den Kapiteln finden Sie auf der folgenden Autoren-Website: <http://www.norusis.com>

Inhalt

1 Einführung in Advanced Statistics 1

2 GLM - Multivariat 2

GLM - Multivariat: Modell	4
Terme konstruieren	5
Quadratsumme	5
Multivariat: Kontraste	7
Kontrasttypen	7
GLM: Multivariat: Profilplots	8
GLM - Multivariat: Post-Hoc-Vergleiche	9
GLM: Speichern	11
Multivariat: Optionen	13
Zusätzliche Funktionen beim Befehl GLM	14

3 GLM - Messwiederholungen 16

Messwertwiederholungen: Faktor(en) definieren	20
Messwiederholungen: Modell	21
Terme konstruieren	22
Quadratsumme	22
Messwiederholungen: Kontraste	23
Kontrasttypen	24
GLM - Messwiederholung: Profilplots	25
GLM - Messwiederholung: Post-Hoc-Mehrfachvergleiche	26
Messwiederholungen: Speichern	28
Messwiederholungen: Optionen	29
Zusätzliche Funktionen beim Befehl GLM	31

4 Analyse der Varianzkomponenten 32

Varianzkomponenten: Modell	34
Terme konstruieren	34

Varianzkomponenten: Optionen	35
Quadratsummen (Varianzkomponenten)	36
Varianzkomponenten: In neuer Datei speichern	37
Zusätzliche Funktionen beim Befehl VARCOMP	37

5 Lineare gemischte Modelle 38

Lineare gemischte Modelle: Variablen für Subjekte und Wiederholungen auswählen	40
Lineare gemischte Modelle: Feste Effekte	42
Nicht verschachtelte Terme konstruieren	42
Verschachtelte Terme konstruieren	43
Quadratsumme	43
Lineare gemischte Modelle: Zufällige Effekte	44
Lineare gemischte Modelle: Schätzung	46
Lineare gemischte Modelle: Statistik	47
Lineare gemischte Modelle: Geschätzte Randmittel	49
Lineare gemischte Modelle: Speichern	50
Zusätzliche Funktionen beim Befehl MIXED	50

6 Generalized Linear Models 52

Verallgemeinerte lineare Modelle - Antwort	57
Verallgemeinerte lineare Modelle – Referenzkategorie	58
Verallgemeinerte lineare Modelle – Einflussvariablen	59
Verallgemeinerte lineare Modelle – Optionen	60
Verallgemeinerte lineare Modelle – Modell	61
Verallgemeinerte lineare Modelle – Schätzung	63
Verallgemeinerte lineare Modelle – Anfangswerte	65
Verallgemeinerte lineare Modelle – Statistiken	66
Verallgemeinerte lineare Modelle – Geschätzte Randmittel	68
Verallgemeinerte lineare Modelle – Speichern	70
Verallgemeinerte lineare Modelle – Exportieren	72
Zusätzliche Funktionen beim Befehl GENLIN	74

7 Generalized Estimating Equations 75

Verallgemeinerte Schätzungsgleichungen – Typ des Modells	79
Verallgemeinerte Schätzungsgleichungen – Antwort	83
Verallgemeinerte Schätzungsgleichungen – Referenzkategorie	84
Verallgemeinerte Schätzungsgleichungen – Einflussvariablen	85
Verallgemeinerte Schätzungsgleichungen – Optionen	86
Verallgemeinerte Schätzungsgleichungen – Modell	87
Verallgemeinerte Schätzungsgleichungen – Schätzung	89
Verallgemeinerte Schätzungsgleichungen – Anfangswerte	91
Verallgemeinerte Schätzungsgleichungen – Statistiken	93
Verallgemeinerte Schätzungsgleichungen – Geschätzte Randmittel	95
Verallgemeinerte Schätzungsgleichungen – Speichern.	97
Verallgemeinerte Schätzungsgleichungen – Exportieren.	99
Zusätzliche Funktionen beim Befehl GENLIN.	100

8 Verallgemeinerte lineare gemischte Modelle 102

Erzeugen eines verallgemeinerten linearen gemischten Modells	104
Ziel	105
Feste Effekte	108
Benutzerdefinierten Term hinzufügen	110
Zufällige Effekte	111
Block für zufällige Effekte	112
Gewichtung und Offset	114
Erstellungsoptionen	115
Geschätzte Mittelwerte	116
Speichern	118
Modellansicht	119
Modellzusammenfassung	119
Datenstruktur	120
Vorhergesagt/Beobachtet	120
Klassifikation	120
Feste Effekte	121
Feste Koeffizienten	121
Kovarianzen der Zufallseffekte.	122
Kovarianz-Parameter	123
Geschätzte Mittelwerte: Signifikante Effekte	123
Geschätzte Mittelwerte: Benutzerdefinierte Effekte	124

9 Modell für loglineare Analyse auswählen 126

Loglineare Analyse: Bereich definieren	127
Loglineare Analyse: Modell	128
Terme konstruieren	128
Modellauswahl für loglineare Analyse: Optionen	129
Zusätzliche Funktionen beim Befehl HILOGLINEAR	130

10 Allgemeine loglineare Analyse 131

Allgemeine loglineare Analyse: Modell	133
Terme konstruieren	133
Allgemeine loglineare Analyse: Optionen	134
Allgemeine loglineare Analyse: Speichern	135
Zusätzliche Funktionen beim Befehl GENLOG	135

11 Logit-loglineare Analyse 137

Logit-loglineare Analyse: Modell	139
Terme konstruieren	140
Logit-loglineare Analyse: Optionen	140
Logit-loglineare Analyse: Speichern	141
Zusätzliche Funktionen beim Befehl GENLOG	142

12 Sterbetafeln (Life Tables Options) 143

Sterbetafeln: Ereignis für Statusvariable definieren	145
Sterbetafeln: Bereich für Faktorvariable definieren	146
Sterbetafeln: Optionen	146
Zusätzliche Funktionen beim Befehl SURVIVAL	147

13 Überlebensanalyse nach Kaplan-Meier 148

Kaplan-Meier: Ereignis für Statusvariable definieren	150
--	-----

Kaplan-Meier: Faktorstufen vergleichen	150
Kaplan-Meier: Neue Variablen speichern	151
Kaplan-Meier: Optionen	152
Zusätzliche Funktionen beim Befehl KM	153

14 Cox-Regressionsanalyse 154

Cox-Regression: Kategoriale Kovariaten definieren	156
Diagramme für Cox-Regressionen	157
Cox-Regression: Neue Variablen speichern	158
Cox-Regression: Optionen	159
Cox-Regression: Ereignis für Statusvariable definieren.	160
Zusätzliche Funktionen beim Befehl COXREG	160

15 Berechnen von zeitabhängigen Kovariaten 161

Berechnen einer zeitabhängigen Kovariate.	162
Zusätzliche Funktionen bei der Cox-Regression mit zeitabhängigen Kovariaten	162

Anhänge

A Kodierungsschemata für kategoriale Variablen 163

Abweichung	163
Einfach.	164
Helmert	165
Differenz.	165
Polynomial	166
Wiederholt	166
Speziell (GLM Ptable)	167
Indikator.	168

<i>B</i>	<i>Kovarianzstrukturen</i>	169
<i>C</i>	<i>Notices</i>	173
	<i>Index</i>	175

Einführung in Advanced Statistics

Die Option “Advanced Statistics” enthält Prozeduren, die komplexere Modellierungsoptionen bieten, als über die Statistics Base-Option 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 nichtkonstante 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” (GLZM) 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 Einflussvariablen 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 Einflussvariablen 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 Wechselwirkungen zwischen den Faktoren als auch die Auswirkungen einzelner Faktoren untersuchen. Außerdem können Sie die Auswirkungen von Kovariaten und Wechselwirkungen zwischen Kovariaten und Faktoren berücksichtigen. Bei der Regressionsanalyse werden die unabhängigen Variablen (Einflußvariablen) als Kovariaten angegeben.

Es können sowohl ausgeglichene als auch nicht ausgeglichene Modelle getestet werden. Ein Design ist ausgeglichen, 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 Matrix-Form 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 Schätzer für die vorhergesagten Mittelwerte der Zellen im Modell und mit Profilplots (Wechselwirkungsdiagrammen) 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 Residuen-Kovarianzmatrix (die SSCP-Matrix für Residuen geteilt durch die Freiheitsgrade der Residuen) und die Residuen-Korrelationsmatrix (die standardisierte Form der Residuen-Kovarianzmatrix) zur Verfügung.

Mit der WLS-Gewichtung können Sie eine Variable angeben, um Beobachtungen für eine WLS-Analyse (Weighted Least Squares, deutsch: 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 Lichtdurchlässigkeit. 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 Wechselwirkung 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.

Statistiken. Post-Hoc-Spannweitentests und Mehrfachvergleiche: geringste signifikante Differenz, Bonferroni, Sidak, Scheffé, multiples F nach Ryan-Einot-Gabriel-Welsch, multiple Spannweite nach Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls-Test, ehrlich signifikante Differenz nach Tukey, 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. Der Levene-Test auf Homogenität der Varianzen. Box- M -Test auf Homogenität der Kovarianzmatrizen für die abhängigen Variablen. Bartlett-Test auf Sphärizität.

Diagramme. Streubreite gegen mittleres Niveau, Residuen-Diagramme, Profilplots (Wechselwirkung).

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

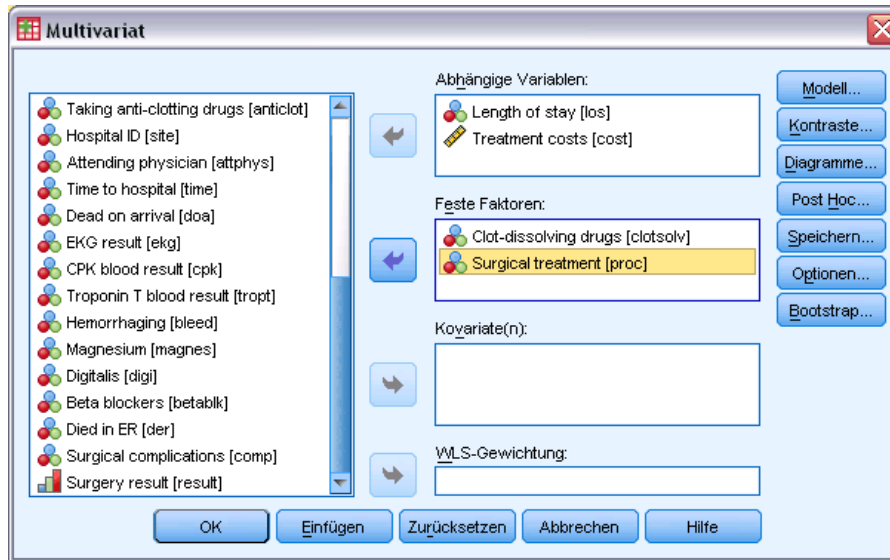
Annahmen. Die Daten für die abhängigen Variablen werden aus einer Zufallsstichprobe von Vektoren aus einer multivariaten, normalverteilten Grundgesamtheit entnommen. In der Grundgesamtheit sind die Varianz-Kovarianz-Matrizen für alle Zellen gleich. Die Varianzanalyse ist unempfindlich gegenüber Abweichungen 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 Residuen-Diagramme 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)

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Allgemeines lineares Modell > Multivariat...

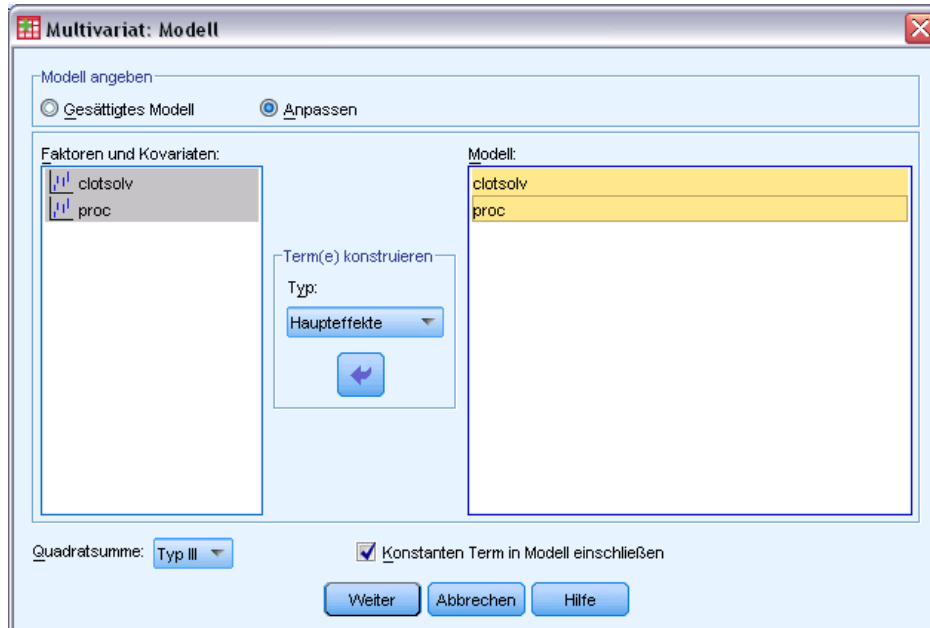
Abbildung 2-1
Dialogfeld "Multivariat"



- Wählen Sie mindestens zwei abhängige Variablen aus.
Wahlweise können Sie feste Faktoren, Kovariaten und eine WLS-Gewichtung angeben.

GLM - Multivariat: Modell

Abbildung 2-2
Dialogfeld "Multivariat: Modell"



Modell angeben. Ein gesättigtes Modell enthält alle Faktoren-Haupteffekte, alle Kovariaten-Haupteffekte und alle faktorweisen Wechselwirkungen. Es enthält keine Kovariaten-Wechselwirkungen. Wählen Sie Anpassen aus, um nur eine Teilmenge von Wechselwirkungen oder Wechselwirkungen 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 Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind.

Quadratsumme. Hier wird die Methode zum Berechnen der Quadratsumme festgelegt. Für ausgeglichene und unausgeglichene 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.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

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 ausgeglichenes ANOVA-Modell, in dem alle Haupteffekte vor den Wechselwirkungseffekten 1. Ordnung festgelegt werden, alle Wechselwirkungseffekte 1. Ordnung wiederum vor den Wechselwirkungseffekten 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 ausgeglichenen ANOVA-Modellen.
- Bei Modellen, die nur Haupteffekte von Faktoren enthalten.
- Bei Regressionsmodellen.
- Bei rein verschachtelten Designs. (Diese Form der Verschachtelung kann durch Verwendung der Befehlssprache erreicht werden.)

Typ III. Voreinstellung. 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 nicht ausgeglichene 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 ausgeglichenen oder unausgeglichenen 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 ausgeglichenen oder unausgeglichenen Modelle mit leeren Zellen.

Multivariat: Kontraste

Abbildung 2-3
Dialogfeld "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 Einheitsmatrix (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 Reihenfolge stehen.

Einfach. Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Diese Art von Kontrast 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 usw. Diese Kontraste werden häufig verwendet, um polynomiale Trends zu schätzen.

GLM: Multivariat: Profilplots

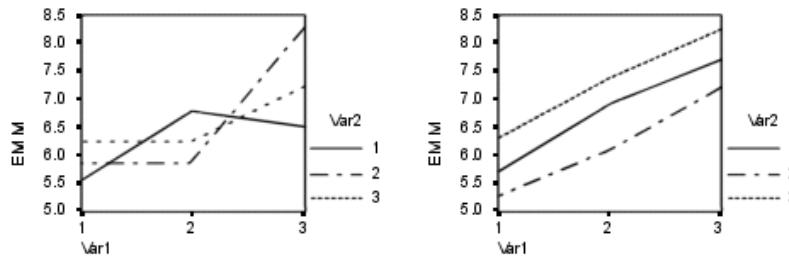
Abbildung 2-4
Dialogfeld "Multivariat: Profilplots"



Profilplots (Wechselwirkungsdiagramme) 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 Wechselwirkung zwischen den Faktoren gibt. Das heißt, dass Sie die Faktorstufen eines einzelnen Faktors untersuchen können. Nichtparallele Linien deuten auf eine Wechselwirkung hin.

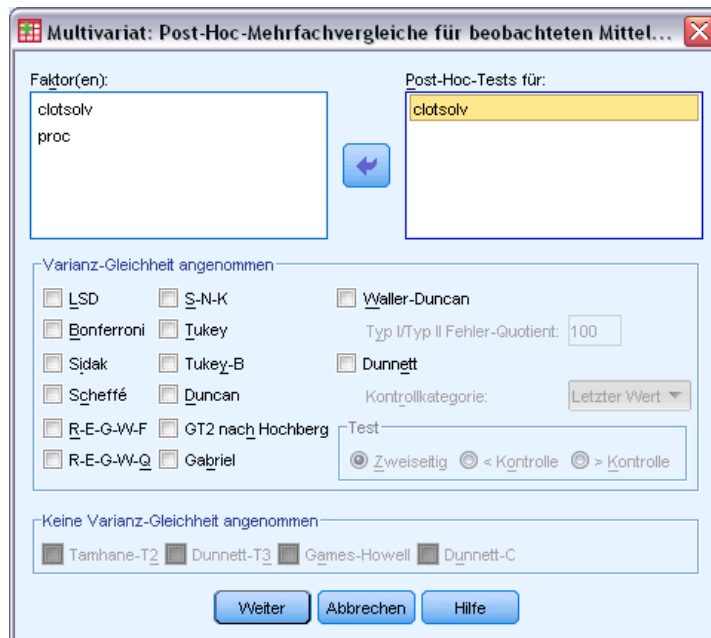
Abbildung 2-5
Nichtparalleles Diagramm (links) und paralleles Diagramm (rechts)



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

Abbildung 2-6
Dialogfeld "Multivariat: Post-Hoc-Mehrfachvergleiche für beobachteten Mittelwert"



Tests für Post-Hoc-Mehrfachvergleiche. Sobald Sie festgestellt haben, daß es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-Hoc-Spannweiten-Tests 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 die ehrlich signifikante Differenz nach Tukey. 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. Die **ehrllich signifikante Differenz nach Tukey** 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 Test auf ehrlich signifikante Differenz nach Tukey leistungsfähiger als der Bonferroni-Test. Bei einer kleinen Anzahl von Paaren ist der Bonferroni-Test leistungsfähiger.

GT2 nach Hochberg ähnelt dem Test auf ehrlich signifikante Differenz nach Tukey, es wird jedoch das studentisierte Maximalmodul verwendet. Meistens ist der Test nach Tukey leistungsfähiger. Der **paarweise Vergleichstest nach Gabriel** verwendet ebenfalls das studentisierte Maximalmodul und zeigt meistens eine größere Schärfe als das GT2 nach Hochberg, wenn die Zellengrößen ungleich sind. Der Test nach Gabriel kann ungenau sein, wenn die Zellengrößen große Abweichungen aufweisen.

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. Wählen Sie $<$ Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors kleiner als der Mittelwert der Kontrollkategorie ist. Wählen Sie $>$ Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors größer als der Mittelwert bei der Kontrollkategorie ist.

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 Teilmengen 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 das **Tamhane-T2** (konservativer paarweiser Vergleichstest auf der Grundlage eines T -Tests), **Dunnett-T3** (paarweiser Vergleichstest auf der Grundlage des studentisierten Maximalmoduls), 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 Mittelwerte von Gruppen 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.

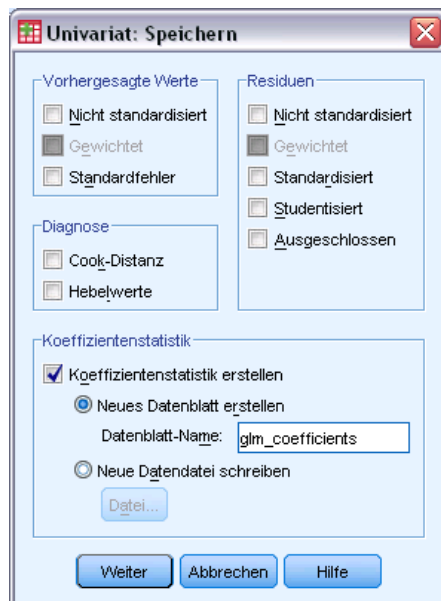
Das Signifikanzniveau des **Scheffé**-Tests ist so festgelegt, dass alle möglichen linearen Kombinationen von Gruppenmittelwerten getestet werden können und nicht nur paarweise Vergleiche verfügbar sind, wie bei dieser Funktion der Fall. Das führt dazu, dass der Scheffé-Test oftmals konservativer als 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 Untergruppen 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. Die ehrlich signifikante Differenz nach Tukey, das GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind multiple Vergleiche, zugleich aber auch Spannweitentests.

GLM: Speichern

Abbildung 2-7
"Speichern"



Vom Modell vorhergesagte Werte, Residuen und verwandte Maße können als neue Variablen im Daten-Editor 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 jeden Fall vorhersagt.

- **Nicht standardisiert (Discriminant Analysis).** Der Wert, den das Modell für die abhängige Variable vorhersagt.
- **Gewichtet.** Gewichtete nichtstandardisierte vorhergesagte Werte. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- **Standardfehler.** Ein Schätzer der Standardabweichung des Durchschnittswerts 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 der unabhängigen Variablen und von Fällen, die einen großen Einfluß auf das Modell haben könnten.

- **Cook-Distanz.** Ein Maß dafür, wie stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von 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.

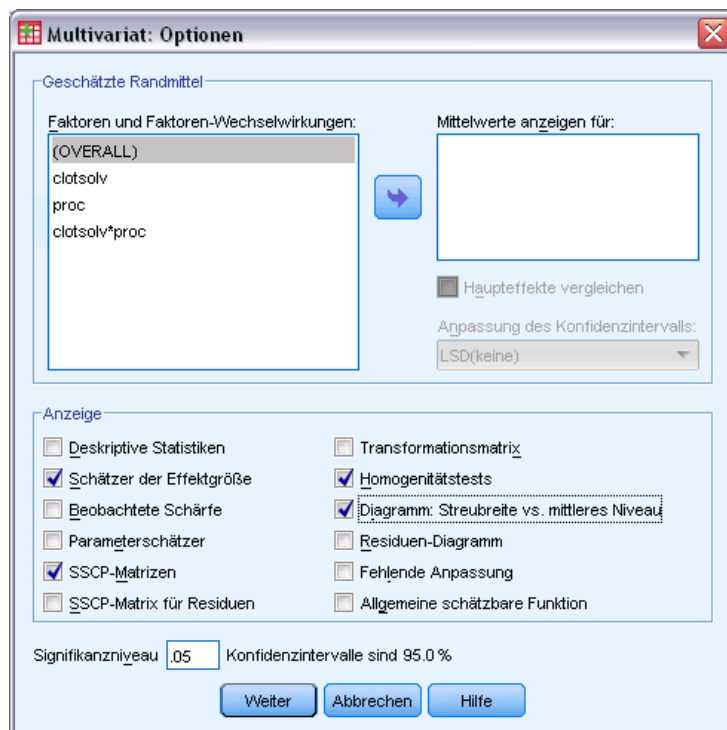
Residuen. Ein nicht standardisiertes Residuum ist der tatsächliche Wert der abhängigen Variablen minus des vom Modell geschätzten Werts. Ebenfalls verfügbar sind standardisierte, studentisierte und ausgeschlossene Residuen. Falls Sie eine WLS-Variable ausgewählt haben, sind auch gewichtete nicht standardisierte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

- **Nicht standardisiert (Discriminant Analysis).** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- **Gewichtet.** Gewichtete nichtstandardisierte Residuen. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- **Standardisiert (Residuals).** Der Quotient aus dem Residuum und einem Schätzer seiner Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Studentisiert.** Ein Residuum, das durch seine geschätzte Standardabweichung geteilt wird, die je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falles und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert.
- **Löschen (Select Cases).** Das Residuum für einen Fall, wenn dieser Fall nicht in die Berechnung der Regressionskoeffizienten eingegangen ist. Es 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 Daten-Set in der aktuellen Sitzung oder in eine externe Datei im SPSS Statistics-Format geschrieben. 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 Matrixdatei auch in anderen Prozeduren verwenden, die Matrixdateien einlesen.

Multivariat: Optionen

Abbildung 2-8
Dialogfeld "Multivariat: Optionen"



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

Geschätzte Randmittel. Wählen Sie die Faktoren und Wechselwirkungen aus, für die Sie Schätzer für die Randmittel der Grundgesamtheit in den Zellen wünschen. Diese Mittel werden gegebenenfalls an die Kovariaten angepaßt. Wechselwirkungen sind nur verfügbar, falls Sie ein benutzerdefiniertes Modell festgelegt haben.

- **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.
- **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 ausgewählt ist.

Anzeigen. Mit der Option Deskriptive Statistik lassen Sie beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen berechnen. Die Option Schätzer 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 Schärfe liefert die Testschärfe, wenn die alternative Hypothese auf die Basis der beobachteten

Werte eingestellt wurde. Mit Parameterschätzer werden Parameterschätzer, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Schärfe 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 Zwischensubjektfaktoren durchgeführt (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. Die Optionen für Diagramme der Streubreite gegen das mittlere Niveau und Residuen-Diagramme sind beim Überprüfen von Annahmen über die Daten nützlich. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind. Wählen Sie Residuen-Diagramme, wenn Sie für jede abhängige Variable ein Residuen-Diagramm (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 ermöglicht Ihnen, einen benutzerdefinierten Hypothesentest zu entwickeln, dessen Grundlage die allgemeine schätzbare Funktion ist. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemeinen schätzbaren Funktion.

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 Schärfe 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. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- 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 benutzerdefinierte fehlende Werte aufnehmen.
- Mit dem Unterbefehl `CRITERIA` können Sie EPS-Kriterien angeben.
- Mit den Unterbefehlen `LMATRIX`, `MMATRIX` und `KMATRIX` können Sie benutzerdefinierte **L**-Matrizen, **M**-Matrizen und **K**-Matrizen erstellen.
- 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 Faktorenwechselwirkungen 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 Matrix-Datendatei erstellen, die Statistiken aus der Zwischensubjekt-ANOVA-Tabelle enthält.
- Mit dem Unterbefehl `OUTFILE` können Sie die Design-Matrix in einer neuen Datendatei speichern.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

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 Wechselwirkungen zwischen den Faktoren als auch die Auswirkungen einzelner Faktoren untersuchen. Zusätzlich können die Effekte von konstanten Kovariaten und die Wechselwirkungen von Kovariaten mit den Zwischensubjektfaktoren berücksichtigt werden.

In einem Design mit doppelt-multivariaten Messwiederholungen stellen die abhängigen Variablen Messungen von mehr als einer 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 ausgeglichene als auch nicht ausgeglichene Modelle getestet werden. Ein Design ist ausgeglichen, 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 Matrix-Form 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 Schätzer für die vorhergesagten Mittelwerte der Zellen im Modell und mit Profilplots (Wechselwirkungsdiagrammen) 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 Residuen-Kovarianzmatrix (die SSCP-Matrix für Residuen geteilt durch die Freiheitsgrade der Residuen) und die Residuen-Korrelationsmatrix (die standardisierte Form der Residuen-Kovarianzmatrix) zur Verfügung.

Mit der WLS-Gewichtung können Sie eine Variable angeben, um Beobachtungen für eine WLS-Analyse (Weighted Least Squares, deutsch: 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 Ergebnisse 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 Wechselwirkung 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 (für Zwischensubjektfaktoren): geringste signifikante Differenz, Bonferroni, Sidak, Scheffé, multiples F nach Ryan-Einot-Gabriel-Welsch, multiple Spannweite nach Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls-Test, ehrlich signifikante Differenz nach Tukey, 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 Häufigkeiten aller abhängigen Variablen in allen Zellen. Levene-Test auf Homogenität der Varianzen. Box- M , Mauchly-Test auf Sphärizität.

Diagramme. Streubreite gegen mittleres Niveau, Residuen-Diagramme, Profilplots (Wechselwirkung).

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 String-Werte besitzen. Innersubjektfaktoren werden im Dialogfeld "Messwertwiederholungen: 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 einen Satz Variablen enthalten. Jeder Satz 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 Nenner-Freiheitsgrade vorgenommen werden, um die univariate F -Statistik zu bestätigen. In der Prozedur "GLM - Messwiederholungen" sind drei Schätzer für diese auch **Epsilon** genannte Anpassung vorhanden. Sowohl die Zähler- als auch die Nenner-Freiheitsgrade 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 gepaarten Stichproben" verwenden.

Ermittlung von "GLM - Messwiederholung"

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Allgemeines lineares Modell > Messwiederholungen...

Abbildung 3-1
Dialogfeld "Messwertwiederholungen: Faktor(en) definieren"

Messwiederholung: Faktor(en) definieren

Name des Innersubjektfaktors:

Anzahl der Stufen:

Hinzufügen

Ändern

Entfernen

time(5)

Messwertbezeichnung:

Hinzufügen

Ändern

Entfernen

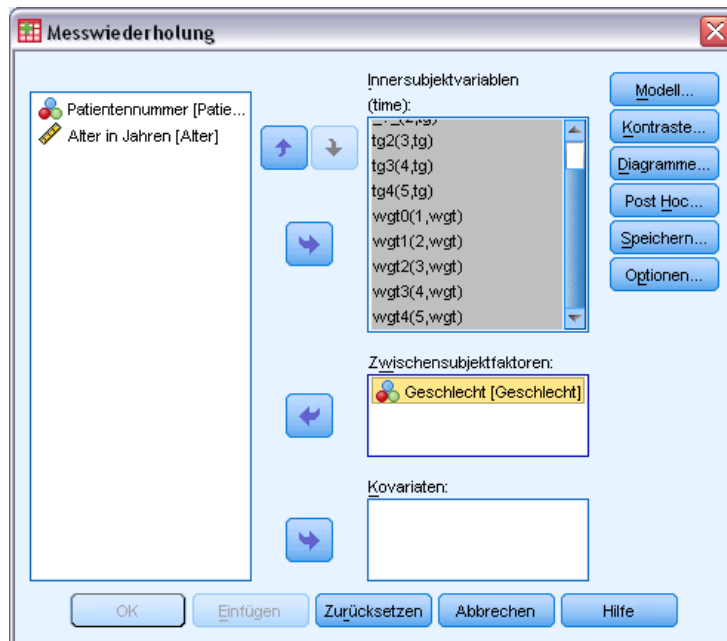
tg

wgt

Definieren Zurücksetzen Abbrechen Hilfe

- ▶ Geben Sie den Namen eines Innersubjektfaktors und die Anzahl seiner Stufen ein.
- ▶ Klicken Sie auf Hinzufügen.
- ▶ Wiederholen Sie diesen Vorgang für jeden Innersubjektfaktor.
So definieren Sie Messwertfaktoren für ein Design mit doppelt-multivariaten Messwiederholungen:
- ▶ Geben Sie die Messwertbezeichnung ein.
- ▶ Klicken Sie auf Hinzufügen.
Nachdem Sie alle Faktoren und Messwerte definiert haben:
- ▶ Klicken Sie auf Definieren.

Abbildung 3-2
Dialogfeld "Messwiederholungen"



- 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 "Messwertwiederholungen: Faktor(en) definieren" erneut öffnen. Hierfür muss das Hauptdialogfeld nicht geschlossen werden. Wahlweise können Sie Zwischensubjektfaktoren und Kovariaten angeben.

Messwertwiederholungen: Faktor(en) definieren

Die Prozedur "GLM - Messwiederholungen" analysiert Gruppen von miteinander in Beziehung stehenden, abhängigen Variablen, die verschiedene Messungen derselben Eigenschaft darstellen. In diesem Dialogfeld können Sie einen oder mehrere Innersubjektfaktoren definieren, die in der Prozedur "GLM - Messwiederholungen" verwendet werden. Unter [Abbildung 3-1](#) auf S. 19. 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.

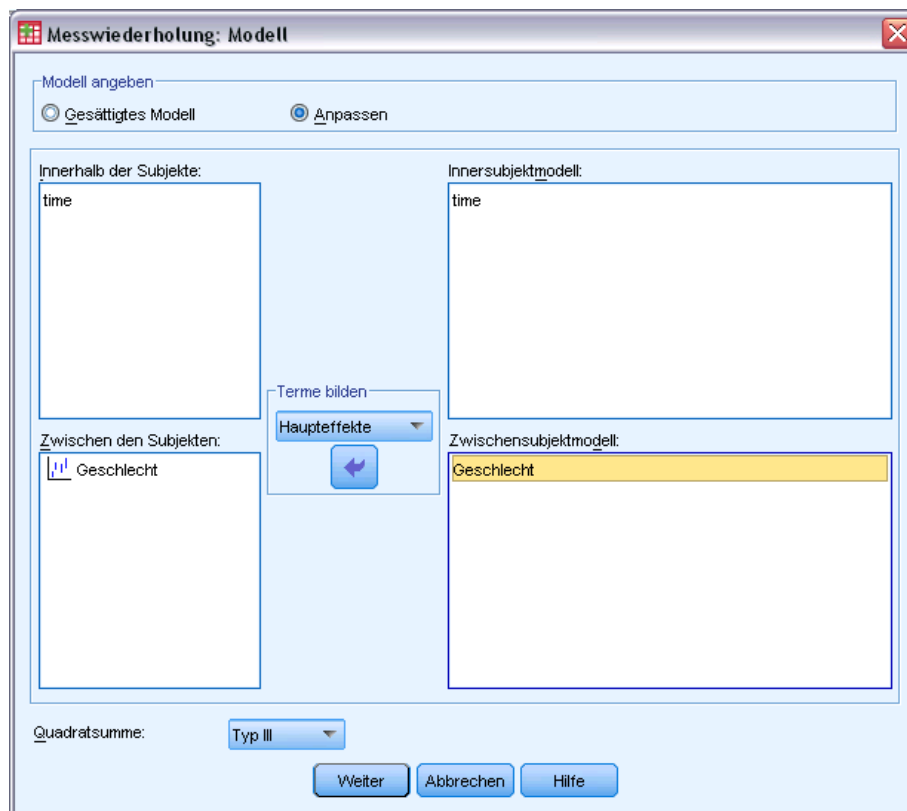
Beispiel. In einer Diätstudie wird das Gewicht von mehreren Personen fünf Wochen lang einmal wöchentlich gemessen. In der Datendatei wird jede Person durch ein Subjekt (also einen Fall) dargestellt. Die Gewichte für die einzelnen Wochen werden in den Variablen *Gewicht1*, *Gewicht2*

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.

Messwerte. 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

Abbildung 3-3
Dialogfeld "Messwiederholungen: Modell"



Modell angeben. Ein gesättigtes Modell enthält alle Faktoren-Haupteffekte, alle Kovariaten-Haupteffekte und alle faktorweisen Wechselwirkungen. Es enthält keine Kovariaten-Wechselwirkungen. Wählen Sie *Anpassen* aus, um nur eine Teilmenge von Wechselwirkungen oder Wechselwirkungen 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 -wechselwirkungen und die Zwischensubjekteffekte und -wechselwirkungen auswählen.

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

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

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 ausgeglichenes ANOVA-Modell, in dem alle Haupteffekte vor den Wechselwirkungseffekten 1. Ordnung festgelegt werden, alle Wechselwirkungseffekte 1. Ordnung wiederum vor den Wechselwirkungseffekten 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 ausgeglichenen ANOVA-Modellen.
- Bei Modellen, die nur Haupteffekte von Faktoren enthalten.
- Bei Regressionsmodellen.
- Bei rein verschachtelten Designs. (Diese Form der Verschachtelung kann durch Verwendung der Befehlssprache erreicht werden.)

Typ III. Voreinstellung. 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 nicht ausgeglichene 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 ausgeglichenen oder unausgeglichenen 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 ausgeglichenen oder unausgeglichenen Modelle mit leeren Zellen.

Messwiederholungen: Kontraste

Abbildung 3-4
Dialogfeld “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 Kontrastkoeffizienten-Matrix, \mathbf{B} der Parametervektor und \mathbf{M} die gemittelte Matrix ist, die der Mittelwert-Transformation 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 Reihenfolge stehen.

Einfach. Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Diese Art von Kontrast 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 usw. Diese Kontraste werden häufig verwendet, um polynomiale Trends zu schätzen.

GLM - Messwiederholung: Profilplots

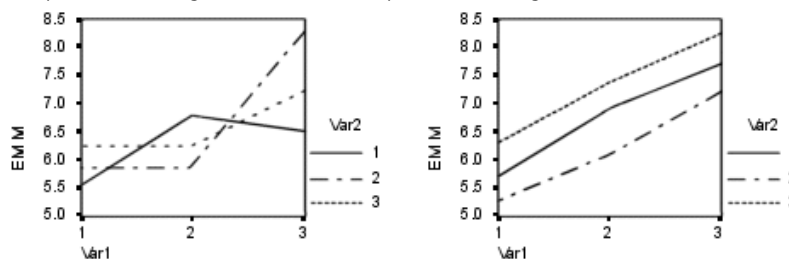
Abbildung 3-5
Dialogfeld "Messwiederholung: Profilplot"



Profilplots (Wechselwirkungsdiagramme) 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 Zwischensubjektfaktoren 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 Wechselwirkung zwischen den Faktoren gibt. Das heißt, dass Sie die Faktorstufen eines einzelnen Faktors untersuchen können. Nichtparallele Linien deuten auf eine Wechselwirkung hin.

Abbildung 3-6
Nichtparalleles Diagramm (links) und paralleles Diagramm (rechts)

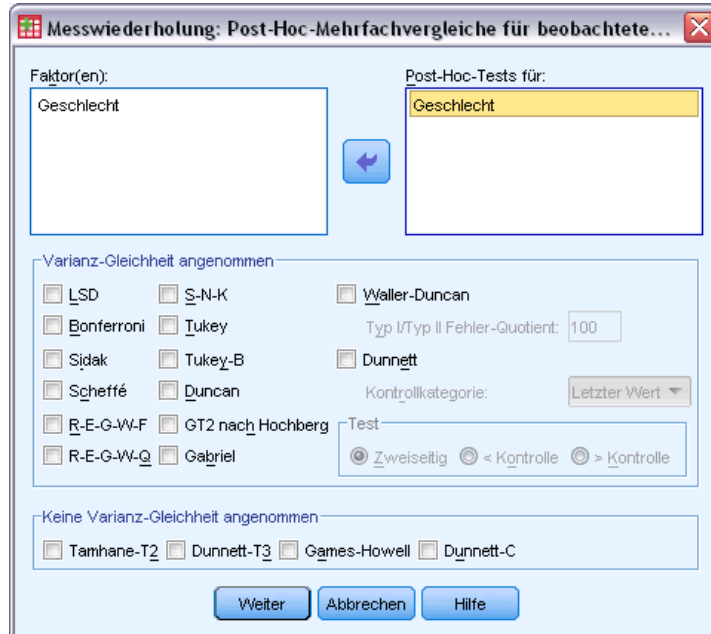


Nachdem ein Diagramm durch Auswahl von Faktoren für die horizontale Achse (und wahlweise von Faktoren für getrennte Linien und getrennte Diagramme) festgelegt wurde, muss das Diagramm der Liste "Diagramme" hinzugefügt werden.

GLM - Messwiederholung: Post-Hoc-Mehrfachvergleiche

Abbildung 3-7

Dialogfeld "Messwiederholung: Post-Hoc-Mehrfachvergleiche für beobachteten Mittelwert"



Tests für Post-Hoc-Mehrfachvergleiche. Sobald Sie festgestellt haben, dass es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-Hoc-Spannweiten-Tests 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 die ehrlich signifikante Differenz nach Tukey. 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. Die **ehrllich signifikante Differenz nach Tukey** 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 Test auf ehrlich signifikante Differenz nach Tukey leistungsfähiger als der Bonferroni-Test. Bei einer kleinen Anzahl von Paaren ist der Bonferroni-Test leistungsfähiger.

GT2 nach Hochberg ähnelt dem Test auf ehrlich signifikante Differenz nach Tukey, es wird jedoch das studentisierte Maximalmodul verwendet. Meistens ist der Test nach Tukey leistungsfähiger. Der **paarweise Vergleichstest nach Gabriel** verwendet ebenfalls das studentisierte Maximalmodul und zeigt meistens eine größere Schärfe als das GT2 nach Hochberg, wenn die Zellengrößen ungleich sind. Der Test nach Gabriel kann ungenau sein, wenn die Zellengrößen große Abweichungen aufweisen.

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. Wählen Sie $<$ Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors kleiner als der Mittelwert der Kontrollkategorie ist. Wählen Sie $>$ Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors größer als der Mittelwert bei der Kontrollkategorie ist.

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 Teilmengen 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 das **Tamhane-T2** (konservativer paarweiser Vergleichstest auf der Grundlage eines T -Tests), **Dunnett-T3** (paarweiser Vergleichstest auf der Grundlage des studentisierten Maximalmoduls), 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 Mittelwerte von Gruppen 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.

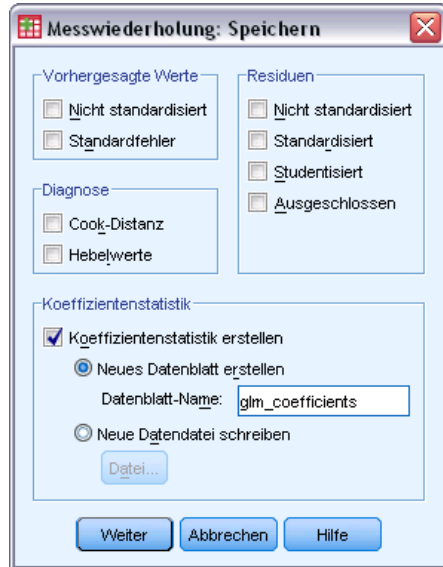
Das Signifikanzniveau des **Scheffé**-Tests ist so festgelegt, dass alle möglichen linearen Kombinationen von Gruppenmittelwerten getestet werden können und nicht nur paarweise Vergleiche verfügbar sind, wie bei dieser Funktion der Fall. Das führt dazu, dass der Scheffé-Test oftmals konservativer als 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 Untergruppen 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. Die ehrlich signifikante Differenz nach Tukey, das GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind multiple Vergleiche, zugleich aber auch Spannweitentests.

Messwiederholungen: Speichern

Abbildung 3-8
Dialogfeld "Messwiederholungen: Speichern"



Vom Modell vorhergesagte Werte, Residuen und verwandte Maße können als neue Variablen im Daten-Editor 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 jeden Fall vorhersagt.

- **Nicht standardisiert (Discriminant Analysis).** Der Wert, den das Modell für die abhängige Variable vorhersagt.
- **Standardfehler.** Ein Schätzer der Standardabweichung des Durchschnittswerts 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 der 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 stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von 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.

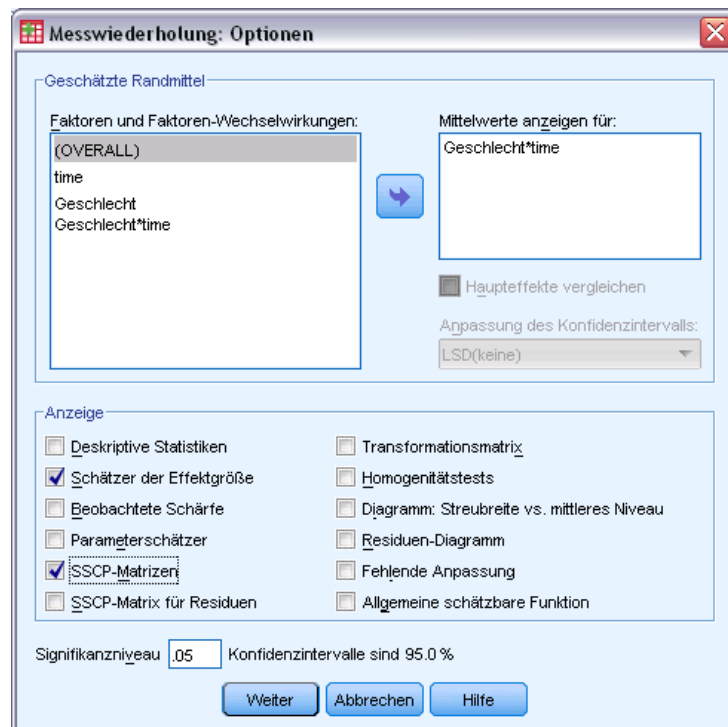
- **Nicht standardisiert (Discriminant Analysis).** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.

- **Standardisiert (Residuals).** Der Quotient aus dem Residuum und einem Schätzer seiner Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Studentisiert.** Ein Residuum, das durch seine geschätzte Standardabweichung geteilt wird, die je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falles und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert.
- **Löschen (Select Cases).** Das Residuum für einen Fall, wenn dieser Fall nicht in die Berechnung der Regressionskoeffizienten eingegangen ist. Es ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.

Koeffizientenstatistik. Hiermit können Sie eine Varianz-Kovarianz-Matrix der Parameterschätzer in ein Daten-Set 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. Daten-Sets 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 Daten-Sets müssen den Regeln zum Benennen von Variablen entsprechen.

Messwiederholungen: Optionen

Abbildung 3-9
Dialogfeld "Messwiederholungen: Optionen"



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

Geschätzte Randmittel. Wählen Sie die Faktoren und Wechselwirkungen aus, für die Sie Schätzer für die Randmittel der Grundgesamtheit in den Zellen wünschen. Diese Mittel werden gegebenenfalls an die Kovariaten angepaßt. Sie können sowohl Innersubjektfaktoren als auch Zwischensubjektfaktoren auswählen.

- **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.
- **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 ausgewählt ist.

Anzeigen. Mit der Option Deskriptive Statistik lassen Sie beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen berechnen. Die Option Schätzer 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 Schärfe liefert die Testschärfe, wenn die alternative Hypothese auf die Basis der beobachteten Werte eingestellt wurde. Mit Parameterschätzer werden Parameterschätzer, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Schärfe 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 Zwischensubjektfaktoren durchgeführt (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. Die Optionen für Diagramme der Streubreite gegen das mittlere Niveau und Residuen-Diagramme sind beim Überprüfen von Annahmen über die Daten nützlich. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind. Wählen Sie Residuen-Diagramme, wenn Sie für jede abhängige Variable ein Residuen-Diagramm (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 ermöglicht Ihnen, einen benutzerdefinierten Hypothesentest zu entwickeln, dessen Grundlage die allgemeine schätzbare Funktion ist. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemeinen schätzbaren Funktion.

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 Schärfe 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. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- 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 benutzerdefinierte fehlende Werte aufnehmen.
- Mit dem Unterbefehl `CRITERIA` können Sie EPS-Kriterien angeben.
- Mit den Unterbefehlen `LMATRIX`, `MMATRIX` und `KMATRIX` können Sie benutzerdefinierte **L**-Matrizen, **M**-Matrizen und **K**-Matrizen erstellen.
- 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 Faktorenwechselwirkungen 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 Matrix-Datendatei erstellen, die Statistiken aus der Zwischensubjekt-ANOVA-Tabelle enthält.
- Mit dem Unterbefehl `OUTFILE` können Sie die Design-Matrix in einer neuen Datendatei speichern.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

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.

Es sind vier verschiedene Methoden zum Schätzen der Varianzkomponenten verfügbar: 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 Schätzer 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 Iterationsprotokoll 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.

Daten. Die abhängige Variable ist quantitativ. Faktoren sind kategorial. Sie können numerische Werte oder String-Werte 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 Residuentermine aus verschiedenen Beobachtungen nicht korrelieren.

Ausgehend von diesen Annahmen sind Beobachtungen aus denselben Stufen eines Zufallsfaktors korreliert. Diese Tatsache unterscheidet ein Varianzkomponenten-Modell 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 Varianzkomponenten-Analyse durchführen. Verwenden Sie zum Testen von Hypothesen die Prozeduren “GLM - Univariat”, “GLM - Multivariat” und “GLM - Messwiederholungen”.

Berechnen von Varianzkomponenten-Analysen

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Allgemeines lineares Modell > Varianzkomponenten...

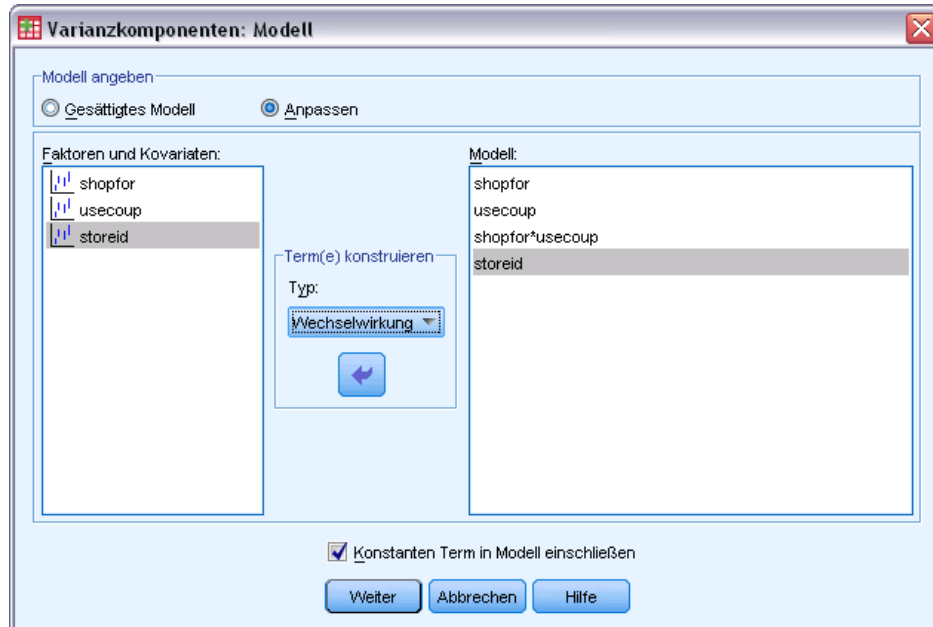
Abbildung 4-1
Dialogfeld “Varianzkomponenten”



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

Varianzkomponenten: Modell

Abbildung 4-2
Dialogfeld "Varianzkomponenten: Modell"



Modell angeben. Ein gesättigtes Modell enthält alle Faktoren-Haupteffekte, alle Kovariaten-Haupteffekte und alle faktorweisen Wechselwirkungen. Es enthält keine Kovariaten-Wechselwirkungen. Wählen Sie Anpassen aus, um nur eine Teilmenge von Wechselwirkungen oder Wechselwirkungen 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 Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind. Das Modell muss einen Zufallsfaktor enthalten.

Konstanten 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.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

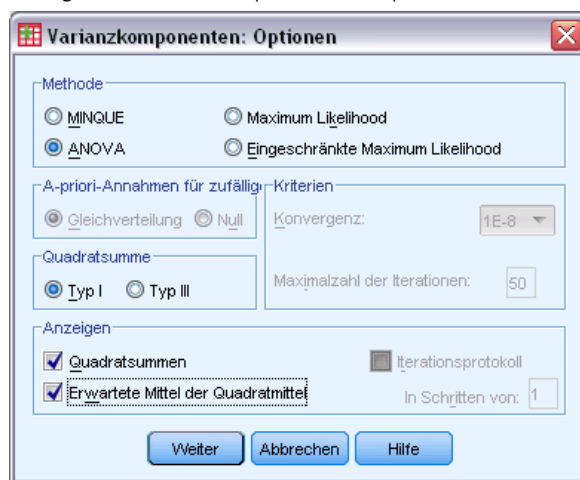
Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Varianzkomponenten: Optionen

Abbildung 4-3
Dialogfeld "Varianzkomponenten: Optionen"



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

- Mit MINQUE (quadratischer unverzerrter Schätzer mit minimaler Norm) werden Schätzer 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 Schätzer unter Verwendung von Quadratsummen vom Typ I oder III für jeden Effekt berechnet. Die ANOVA-Methode liefert manchmal negative Varianz-Schätzer. 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 Schätzer berechnet, die am ehesten den tatsächlich beobachteten Daten entsprechen. Diese Schätzer können jedoch verzerrt sein. Diese Methode ist asymptotisch normalverteilt. ML- und REML-Schätzer sind translationsinvariant. Bei dieser Methode werden die für die Schätzung der festen Effekte verwendeten Freiheitsgrade nicht berücksichtigt.
- Mit Schätzern der eingeschränkten Maximum Likelihood (REML) werden die ANOVA-Schätzer für viele (wenn nicht sogar alle) Fälle von ausgeglichenen Daten reduziert. Da diese Methode für die festen Effekte korrigiert 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 - zufällige Effekte. Gleich geht davon aus, dass alle Zufallseffekte und der Residuenterm den gleichen Einfluß auf die Beobachtung haben. Das Null-Schema entspricht der Annahme von Null-Varianzen 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 Varianz-Schätzer 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 ein Iterationsprotokoll 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 angepaßt. Die Methode mit Quadratsummen vom Typ I wird gewöhnlich in den folgenden Fällen verwendet:

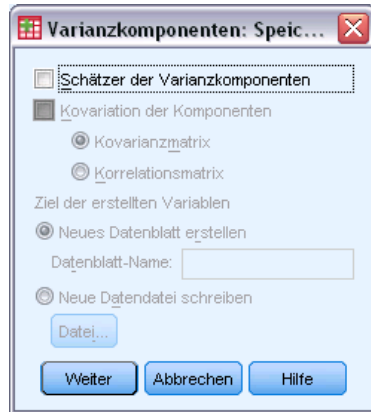
- Ein ausgeglichenes ANOVA-Modell, in dem alle Haupteffekte vor den Wechselwirkungseffekten 1. Ordnung festgelegt werden, alle Wechselwirkungseffekte 1. Ordnung wiederum vor den Wechselwirkungseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme 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. Voreinstellung. 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 nicht ausgeglichene 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 ausgeglichenen oder unausgeglichenen Modelle ohne leere Zellen.

Varianzkomponenten: In neuer Datei speichern

Abbildung 4-4
Dialogfeld "Varianzkomponenten: In neuer Datei speichern"



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

Schätzer der Varianzkomponenten. Hiermit können Sie die Schätzer der Varianzkomponenten und Labels für die Schätzer in einer Datendatei oder in einem Daten-Set 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 Daten-Set 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 Daten-Sets oder einer externen Datei festlegen, welches bzw. welche die Schätzer der Varianzkomponenten und/oder die Matrix enthält. Daten-Sets 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 Daten-Sets 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`

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

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

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Lineare gemischte Modelle

Die Prozedur “Lineare gemischte Modelle” erweitert das allgemeine lineare Modell, indem sie zulässt, dass die Daten korrelierte und nichtkonstante 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 Werkzeug 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 zu korrigieren.

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

Statistiken. Deskriptive Statistiken: Stichprobengrößen, Mittelwerte und Standardabweichungen von abhängigen Variablen und Kovariaten für jede einzelne Kombination von Faktorstufen. Faktorstufeninformationen: sortierte Werte der Stufen für jeden Faktor und ihre Häufigkeiten. Darüber hinaus Parameterschätzer 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.

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

Annahmen. 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.

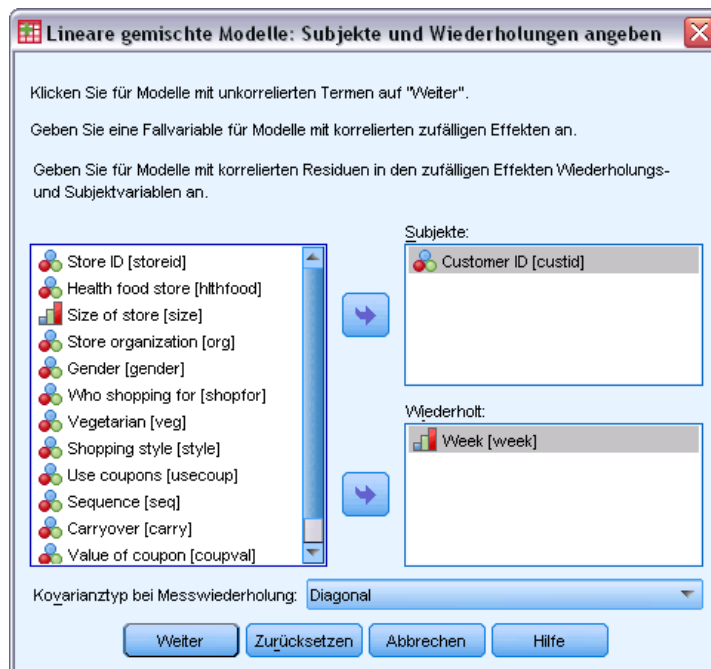
Verwandte Prozeduren. Verwenden Sie die Prozedur “Explorative Datenanalyse”, um die Daten vor der Durchführung der Analyse zu untersuchen. Wenn Sie keine korrelierte oder nichtkonstante Variabilität vermuten, können Sie die Prozedur “GLM - Univariat” oder die Prozedur “GLM - Messwiederholung” verwenden. 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

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Gemischte Modelle > Linear...

Abbildung 5-1

Dialogfeld “Lineare gemischte Modelle: Variablen für Subjekte und Wiederholungen angeben”



- ▶ Wählen Sie optional mindestens eine Subjektvariable aus.
- ▶ Wählen Sie optional mindestens eine Wiederholungsvariable aus.
- ▶ Wählen Sie optional eine Kovarianzstruktur der Residuen aus.
- ▶ Klicken Sie auf Weiter.

Abbildung 5-2
Dialogfeld "Lineare gemischte Modelle"



- ▶ Wählen Sie eine abhängige Variable aus.
- ▶ Wählen Sie mindestens einen Faktor oder eine Kovariante aus.
- ▶ 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: Variablen für Subjekte und Wiederholungen auswählen

In diesem Dialogfeld können Sie Variablen auswählen, die Subjekte und Beobachtungswiederholungen definieren, und eine Kovarianzstruktur für die Residuen angeben. Unter [Abbildung 5-1](#) auf S. 39.

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.

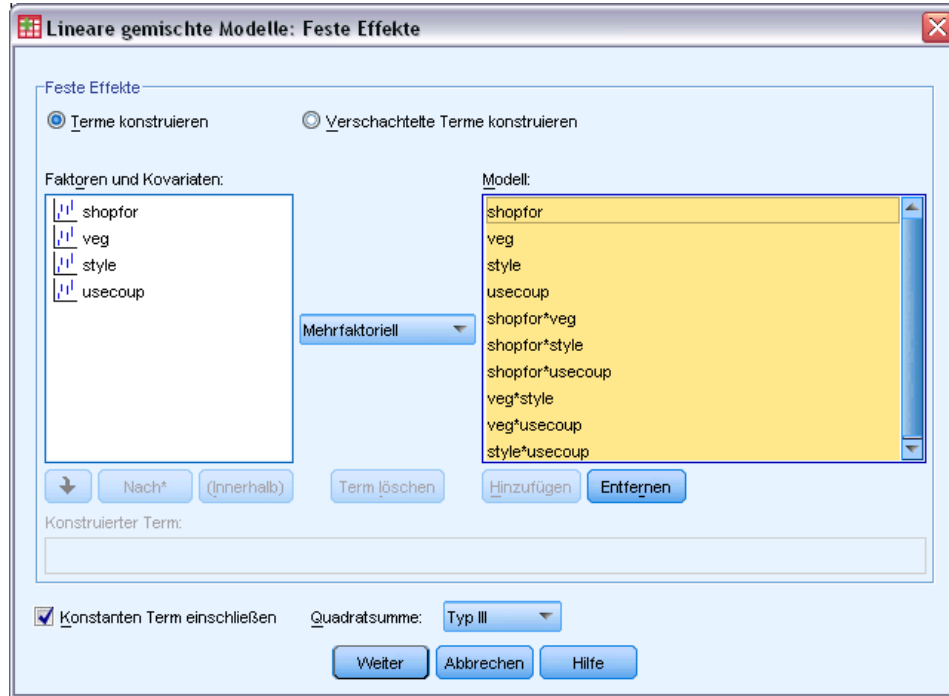
Art der wiederholten Kovarianz. Hiermit wird die Kovarianzstruktur für die Residuen angegeben. Die verfügbaren Strukturen lauten folgendermaßen:

- Ante-Dependenz: 1. Ordnung
- AR(1)
- AR(1): Heterogen
- ARMA(1,1)
- Zusammengesetzt symmetrisch (ZS)
- Zusammengesetzt symmetrisch: Korrelationsmetrik
- Zusammengesetzt symmetrisch: Heterogen
- Diagonal
- Faktor-analytisch: 1. Ordnung
- Faktor-analytisch: 1. Ordnung, heterogen.
- Huynh-Feldt (HF)
- Skalierte Identität
- Toeplitz
- Toeplitz: Heterogen
- Unstrukturiert (UN)
- Unstrukturiert: Korrelationen

Für weitere Informationen siehe Thema Kovarianzstrukturen in Anhang B auf S. 169.

Lineare gemischte Modelle: Feste Effekte

Abbildung 5-3
Dialogfeld "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 konstruieren.

Konstanten Term einschließen. 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.

Quadratsumme. Hier wird die Methode zum Berechnen der Quadratsumme festgelegt. Für Modelle ohne fehlende Zellen wird meistens die Methode vom Typ III verwendet.

Nicht verschachtelte Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Mehrfaktoriell. Hiermit werden alle möglichen Wechselwirkungen und Haupteffekte der ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme konstruieren

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. 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 Wechselwirkungseffekte einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Wechselwirkung 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.

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 ausgeglichenes ANOVA-Modell, in dem alle Haupteffekte vor den Wechselwirkungseffekten 1. Ordnung festgelegt werden, alle Wechselwirkungseffekte 1. Ordnung wiederum vor den Wechselwirkungseffekten 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. Voreinstellung. 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 nicht ausgeglichene 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 ausgeglichenen oder unausgeglichenen Modelle ohne leere Zellen.

Lineare gemischte Modelle: Zufällige Effekte

Abbildung 5-4
Dialogfeld "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 (ZS)
- Zusammengesetzt symmetrisch: Korrelationsmetrik
- Zusammengesetzt symmetrisch: Heterogen
- Diagonal
- Faktor-analytisch: 1. Ordnung
- Faktor-analytisch: 1. Ordnung, heterogen.
- Huynh-Feldt (HF)
- Skalierte Identität
- Toeplitz
- Toeplitz: Heterogen
- Unstrukturiert (UN)
- Unstrukturiert: Korrelationsmetrik
- Varianzkomponenten

Für weitere Informationen siehe [Thema Kovarianzstrukturen in Anhang B auf S. 169](#).

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

Sie können mehrere Modelle mit Zufallseffekten angeben. Klicken Sie nach der Konstruktion des ersten Modells auf *Weiter*, um ein weiteres Modell zu konstruieren. 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.

Lineare gemischte Modelle: Schätzung

Abbildung 5-5
Dialogfeld "Lineare gemischte Modelle: Schätzung"

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

Iterationen:

- **Maximale Anzahl der Iterationen.** Geben Sie eine nichtnegative Ganzzahl an.
- **Maximalzahl für Schritt-Halbierungen.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritt-Halbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationsprotokoll ausgeben alle n Schritte.** Zeigt eine Tabelle an, die den Funktionswert und die Parameterschätzer der Log-Likelihood nach allen n Iterationen enthält, ausgehend von der 0. Iteration (den Anfangsschätzungen). Wenn Sie das Iterationsprotokoll 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.

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

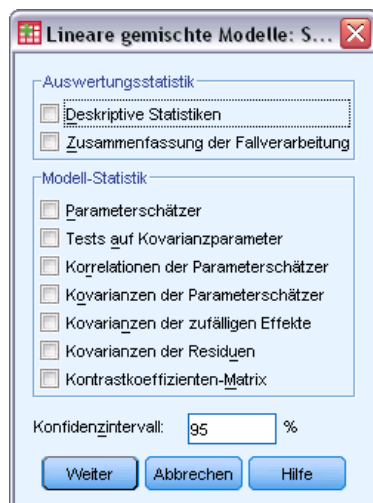
Konvergenz der Hesse-Matrix. Für die Spezifikation Absolut wird angenommen, dass eine Konvergenz vorliegt, wenn eine Statistik auf der Basis der Hesse-Matrix kleiner als der angegebene Wert ist. Für die Spezifikation Relativ wird angenommen, dass eine Konvergenz vorliegt, wenn die Statistik kleiner als 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 Wertungsschritte. Fordert die Verwendung des Fisher-Bewertungsalgorithmus bis zur n -ten Iteration an. Geben Sie eine positive Ganzzahl ein.

Toleranz für Prüfung auf Singularität. Der Wert, der zur Prüfung auf Singularität als Toleranz verwendet wird. Geben Sie einen positiven Wert ein.

Lineare gemischte Modelle: Statistik

Abbildung 5-6
Dialogfeld "Lineare gemischte Modelle: Statistik"



Auswertungsstatistik. Es werden die folgenden Tabellen erstellt:

- **Deskriptive Statistik.** 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.

Modell-Statistik. Es werden die folgenden Tabellen erstellt:

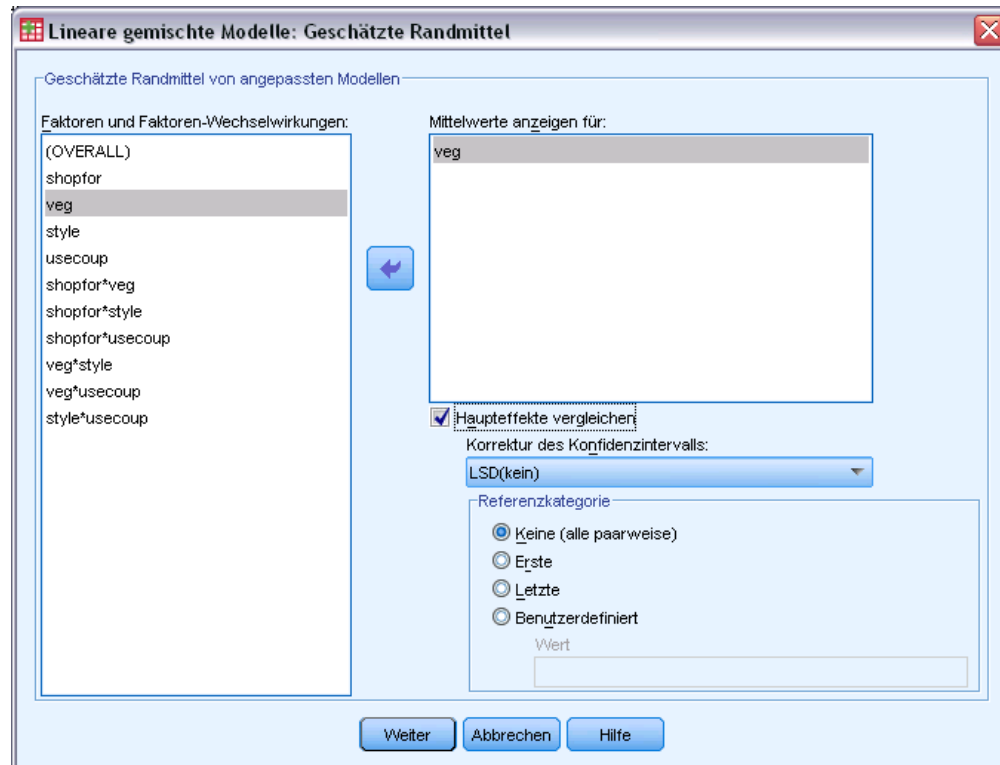
- **Parameterschätzer.** Zeigt die Parameterschätzer mit festen Effekten und Zufallseffekten sowie deren approximativen Standardfehler an.
- **Tests auf Kovarianzparameter.** Zeigt die asymptotischen Standardfehler und Wald-Tests für die Kovarianzparameter an.
- **Korrelationen der Parameterschätzer.** Zeigt die asymptotische Korrelationsmatrix der Parameterschätzer mit festen Effekten an.
- **Kovarianzen der Parameterschätzer.** Zeigt die asymptotische Kovarianzmatrix der Parameterschätzer 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.
- **Matrix Kontrastkoeffizienten.** 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 oder gleich 0 und kleiner als 100 ein. Der Standardwert ist 95.

Lineare gemischte Modelle: Geschätzte Randmittel

Abbildung 5-7

Dialogfeld "Lineare gemischte Modelle: Geschätzte Randmittel"



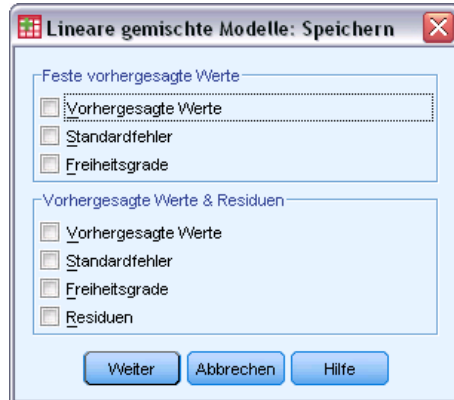
Geschätzte Randmittel von angepassten Modellen. Mit dieser Gruppe können Sie vom Modell vorhergesagte geschätzte Randmittel der abhängigen Variable in den Zellen und die dazugehörigen Standardfehler für die festgelegten Faktoren anfordern. Außerdem können Sie einen Vergleich der Faktorstufen von Haupteffekten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

- **Faktoren und Faktoren-Wechselwirkungen.** Diese Liste enthält Faktoren und Faktoren-Wechselwirkungen, die im Dialogfeld "Fest" angegeben wurden, sowie einen OVERALL-Term. Aus Kovariaten konstruierte Modellterme sind aus dieser Liste ausgeschlossen.
- **Mittelwerte anzeigen für.** Dieses Verfahren berechnet die geschätzten Randmittel für die Faktoren und Faktoren-Wechselwirkungen, 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 Faktoren-Wechselwirkungen 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. Folgende Methoden sind verfügbar: 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

Abbildung 5-8
Dialogfeld "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 Schätzer.
- **Freiheitsgrade.** Die den Schätzern zugeordneten Freiheitsgrade.

Vorhergesagte Werte & Residuen. Speichert Variablen, die sich auf den an das Modell angepassten Wert beziehen.

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

Zusätzliche Funktionen beim Befehl MIXED

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- 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 benutzerdefinierte 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 Wechselwirkungen vergleichen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Generalized Linear Models

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 auf die Anzahl der Havarien für mehrere Schiffstypen anzuwenden, die in verschiedenen Zeiträumen gebaut wurden. Anhand des so entstandenen Modells kann ermittelt werden, welche Schiffstypen am havarieanfälligen sind.

Ein KFZ-Versicherungsunternehmen kann mithilfe von verallgemeinerten linearen Modellen eine Gammaregression an die Schadensansprüche für Autos anpassen. Anhand des so entstandenen Modells können die Faktoren ermittelt werden, die am meisten zur Anspruchshöhe beitragen.

Medizinforscher können mithilfe von verallgemeinerten linearen Modellen eine komplementäre Log-Log-Regression für intervallzensierte Überlebensdaten anpassen, um die Dauer bis zum Wiederauftreten eines Krankheitsbilds vorherzusagen.

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, Skalengewicht und Offset metrisch (Skala) sind.

Annahmen. Fälle werden als unabhängige Beobachtungen betrachtet.

So erstellen Sie ein verallgemeinertes lineares Modell:

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren > Generalized Linear Models > Verallgemeinerte lineare Modelle...

Abbildung 6-1
Verallgemeinerte lineare Modelle - Registerkarte "Typ des Modells"

- ▶ Geben Sie eine Verteilung und eine Verknüpfungsfunktion an (unten finden Sie Einzelheiten zu den verschiedenen Optionen).
- ▶ Wählen Sie auf der Registerkarte [Antwort](#) eine abhängige Variable aus.
- ▶ Wählen Sie auf der Registerkarte [Einflussvariablen](#) die Faktoren und Kovariaten aus, die zur Vorhersage der abhängigen Variablen verwendet werden sollen.
- ▶ Geben Sie auf der Registerkarte [Modell](#) Modelleffekte mithilfe der ausgewählten Faktoren und Kovariaten 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 Antwort.

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

Ordinale Antwort.

- **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.

- **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”.

Binäre Antwort oder Ereignisse/Versuche-Daten.

- **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.

- **Tweedie mit Log-Verknüpfung.** Gibt “Tweedie” als Verteilung und “Log” als Verknüpfungsfunktion an.
- **Tweedie mit Identität-Verknüpfung.** Gibt “Tweedie” als Verteilung und “Identität” als Verknüpfungsfunktion an.

Benutzerdefiniert. Gibt 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 Daten-Set 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 nichtnegativen 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 des Hilfsparameters der negativen Binomialverteilung kann jede beliebige Zahl größer oder gleich 0 sein. Sie können einen festen Wert dafür festlegen oder den Wert durch die Prozedur schätzen lassen. Wenn der Hilfsparameter auf 0 gesetzt wird, entspricht die Verwendung dieser Verteilung der 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.** Diese Verteilung lässt sich als Anzahl der Vorkommnisse eines untersuchten Ereignisses in einem festen Zeitraum betrachten und eignet sich für Variablen mit nichtnegativen ganzzahligen Variablen mit nichtnegativen 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 Gamma-Verteilungen repräsentiert werden können; die Verteilung ist dahingehend “gemischt”, dass sie sowohl Eigenschaften von stetigen Verteilungen (nimmt nichtnegative 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 Antwort repräsentieren. Bei der abhängigen Variablen kann es sich um eine numerische Variable oder eine String-Variable handeln; sie muss mindestens zwei verschiedene gültige Datenwerte aufweisen.

Verknüpfungsfunktionen

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

- **Identität.** $f(x)=x$. Die abhängige Variable wird nicht transformiert. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Log-Log komplementär.** $f(x)=\log(-\log(1-x))$. Nur für die Binomialverteilung geeignet.
- **Cauchit (kumulativ).** $f(x) = \tan(\pi (x - 0,5))$ auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Nur für die Multinomialverteilung geeignet.

- **Log-Log komplementär (kumulativ)**, $f(x) = \ln(-\ln(1-x))$, auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. 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.
- **Log-Log negativ (kumulativ)**, $f(x) = -\ln(-\ln(x))$, auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Nur für die Multinomialverteilung geeignet.
- **Probit (kumulativ)**, $f(x) = \Phi^{-1}(x)$, auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet, wobei Φ^{-1} die umgekehrte kumulative Standard-Normalverteilungsfunktion ist. Nur für die Multinomialverteilung geeignet.
- **Log**, $f(x) = \log(x)$. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Log. Komplement**, $f(x) = \log(1-x)$. Nur für die Binomialverteilung geeignet.
- **Logit**, $f(x) = \log(x / (1-x))$. Nur für die Binomialverteilung geeignet.
- **Negativ binomial**, $f(x) = \log(x / (x+k^{-1}))$, wobei k der Hilfsparameter der negativen Binomialverteilung ist. Nur für die negative Binomialverteilung geeignet.
- **Log-Log negativ**, $f(x) = -\log(-\log(x))$. Nur für die Binomialverteilung geeignet.
- **Odds-Potenz**, $f(x) = [(x/(1-x))^\alpha - 1] / \alpha$, wenn $\alpha \neq 0$. $f(x) = \log(x)$, wenn $\alpha = 0$. α ist die erforderliche Zahlenangabe. Es muss sich dabei um eine reelle Zahl handeln. Nur für die Binomialverteilung geeignet.
- **Probit**, $f(x) = \Phi^{-1}(x)$, wobei Φ^{-1} die umgekehrte kumulative Standard-Normalverteilungsfunktion ist. Nur für die Binomialverteilung geeignet.
- **Exponent**, $f(x) = x^\alpha$, wenn $\alpha \neq 0$. $f(x) = \log(x)$, wenn $\alpha = 0$. α ist die erforderliche Zahlenangabe. Es muss sich dabei um eine reelle Zahl handeln. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.

Verallgemeinerte lineare Modelle - Antwort

Abbildung 6-2
Dialogfeld "Verallgemeinerte lineare Modelle"

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 Stringvariablen oder numerische Variablen sein.
- **Anzahl der Ereignisse, die in einer Menge von Versuchen eintreten.** Wenn die Antwort eine Anzahl von Ereignissen ist, die in einer Menge von Versuchen eintreten, 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.

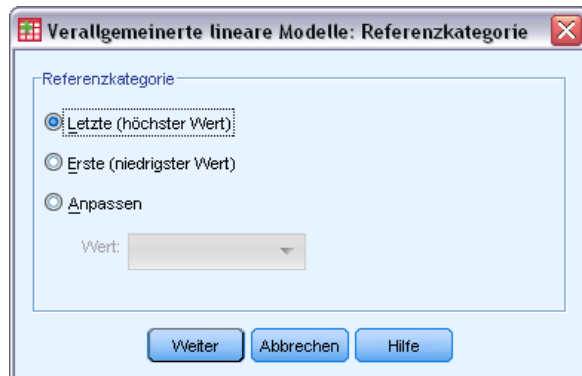
Bei den Ereignissen sollte es sich um nichtnegative Ganzzahlen und bei den Versuchen um positive Ganzzahlen handeln.

Bei ordinalen 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).

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

Verallgemeinerte lineare Modelle – Referenzkategorie

Abbildung 6-3
Verallgemeinerte lineare Modelle – Dialogfeld "Referenzkategorie"

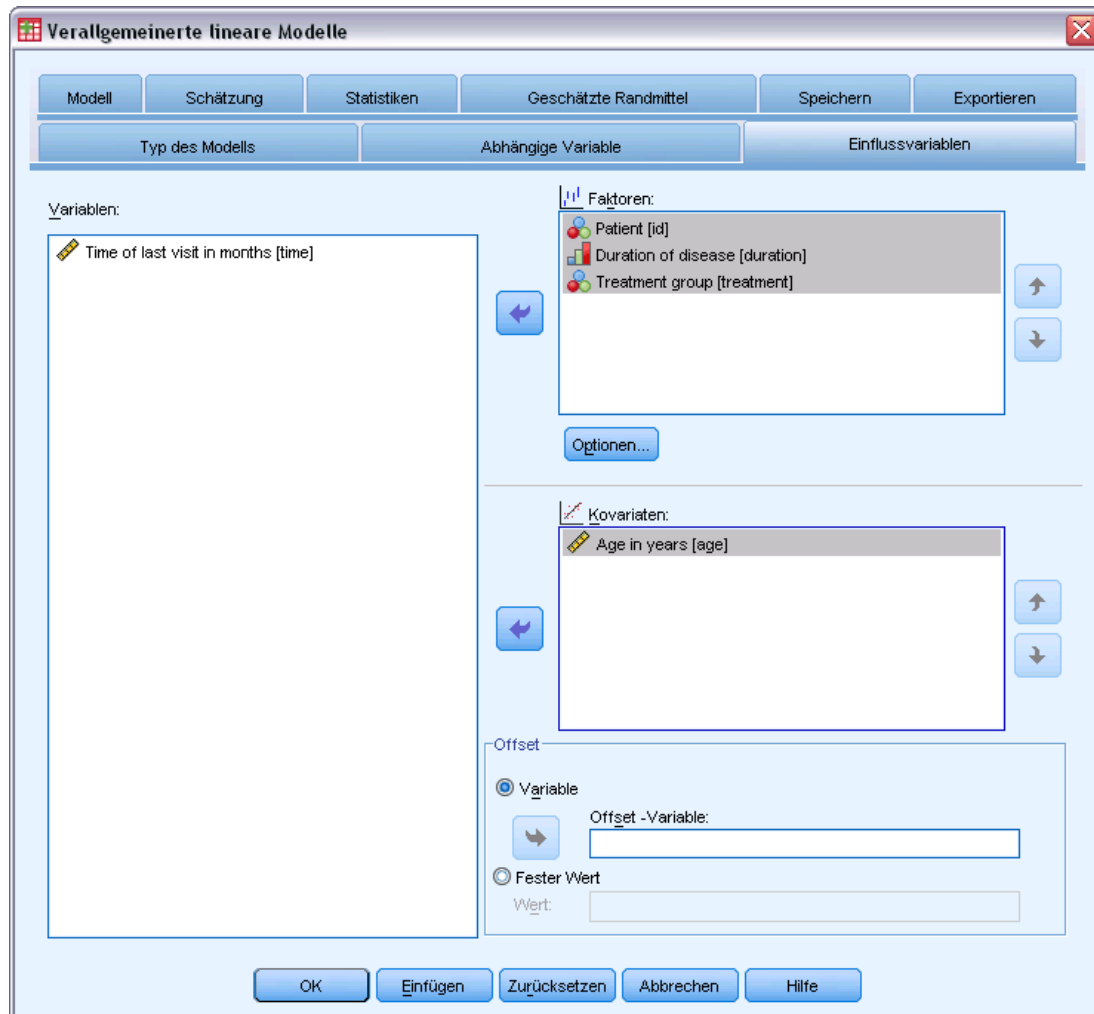


Bei binären Antworten können Sie die Referenzkategorie für die abhängige Variable auswählen. Dies kann sich auf bestimmte Ausgaben, wie beispielsweise Parameterschätzer und gespeicherte Werte, auswirken, sollte jedoch nicht die Anpassungsgüte des Modells verändern. Beispiel: Angenommen, Ihre binäre Antwort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

Verallgemeinerte lineare Modelle – Einflussvariablen

Abbildung 6-4
Generalized Linear Models: Registerkarte "Einflussvariablen"



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

Faktoren. Faktoren sind kategoriale Einflussvariablen; es kann sich dabei um numerische Variablen oder Stringvariablen handeln.

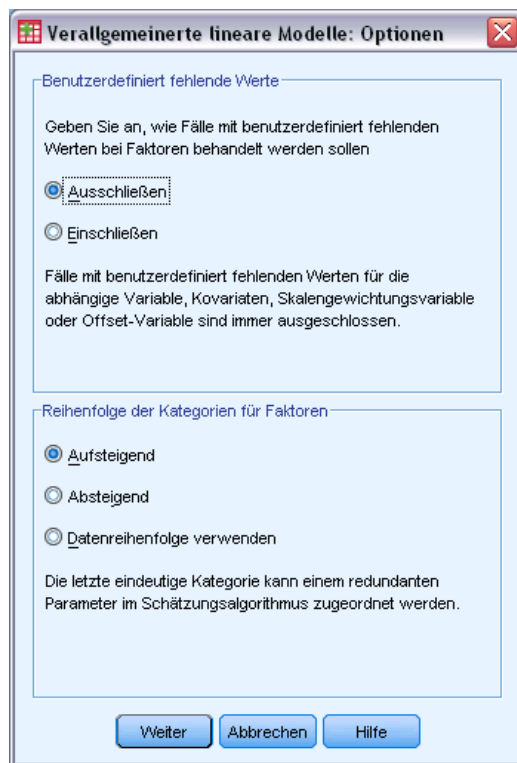
Kovariaten. Kovariaten sind metrische Einflussvariablen; sie müssen numerisch sein.

Anmerkung: 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 dieselbe Menge an Einflussvariablen während mehrerer Ausführungen der Prozedur beibehalten, um eine konsistente Anzahl an Teilgesamtheiten zu gewährleisten.

Offset. Der Term “Offset” ist eine “strukturelle” Einflussvariable. Ihr Koeffizient wird nicht vom Modell geschätzt, sondern es wird davon ausgegangen, dass er den Wert 1 aufweist. Daher werden die Werte des Offsets einfach zur linearen Einflussvariablen der abhängigen Variablen 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 poissonverteilte Response-Variable modelliert werden, wenn die Länge der Fahrpraxis als Offset-Term einbezogen wird.

Verallgemeinerte lineare Modelle – Optionen

Abbildung 6-5
Verallgemeinerte lineare Modelle – Dialogfeld “Optionen”



Diese Optionen werden auf alle auf der Registerkarte “Einflussvariablen” angegebenen Faktoren angewendet.

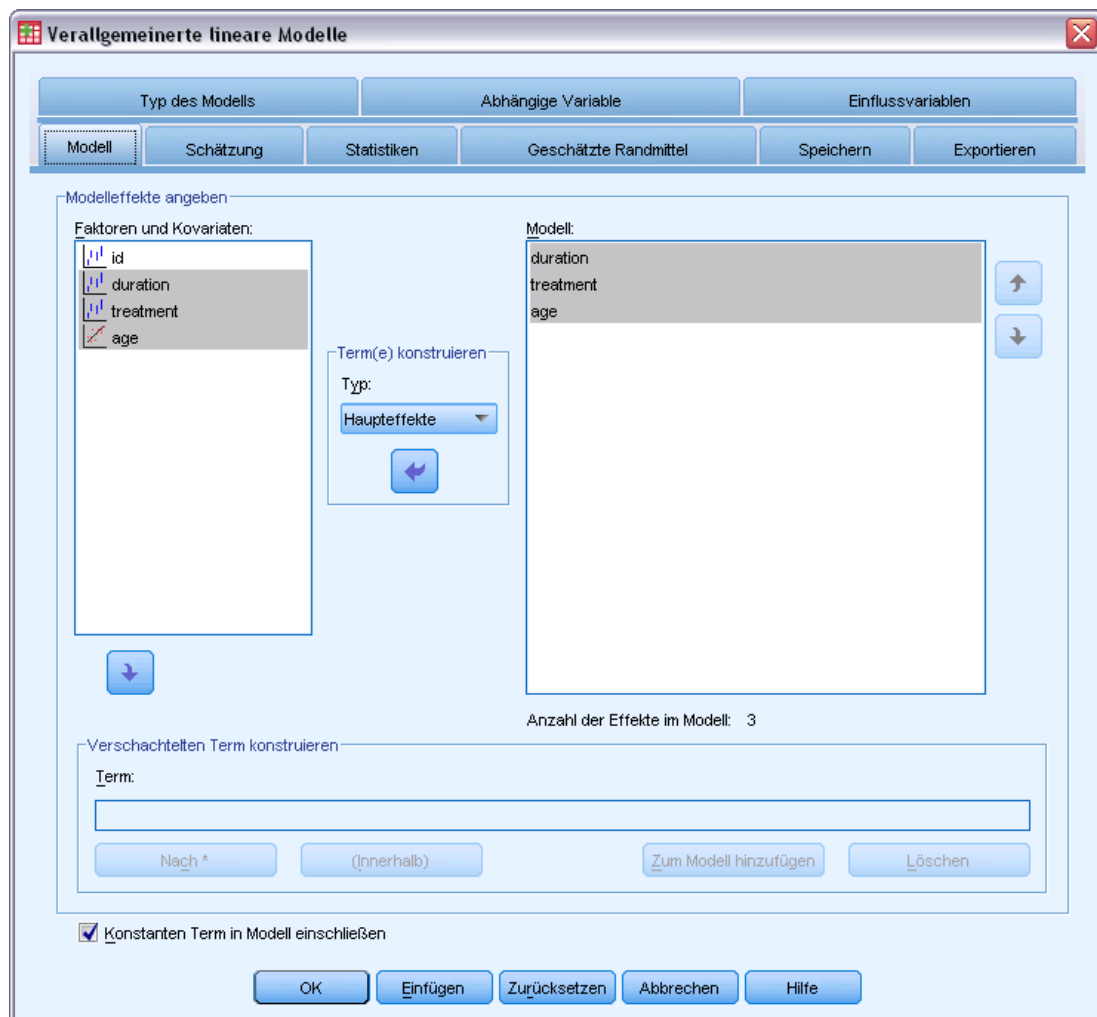
Benutzerdefinierte fehlende Werte. Faktoren müssen gültige Werte für einen Fall aufweisen, um in die Analyse aufgenommen zu werden. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Faktorvariablen als gültige Werte behandelt werden sollen.

Kategorienreihenfolge. Diese ist relevant für die Bestimmung der letzten Stufe eines Faktors, die einem redundanten Parameter im Schätzalgorithmus zugeordnet werden kann. Durch das Ändern der Kategorienreihenfolge können sich die Werte der Faktorstufen-Effekte ändern, da

diese Parameterschätzer relativ zur “letzten” 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

Abbildung 6-6
Generalized Linear Models: Registerkarte “Modell”



Modell-Effekte angeben. Das Standardmodell ist “Nur konstanter Term”. Daher müssen Sie die anderen Modelleffekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme konstruieren.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

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

Mehrfaktoriell. Hiermit werden alle möglichen Wechselwirkungen und Haupteffekte der ausgewählten Variablen erzeugt.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. 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 Wechselwirkungseffekte, 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 Wechselwirkung 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 Schwellenparameter, die Übergangspunkte zwischen benachbarten Kategorien definieren. Die Schwellenwerte werden immer in das Modell aufgenommen.

Verallgemeinerte lineare Modelle – Schätzung

Abbildung 6-7
Generalized Linear Models: Registerkarte "Schätzung"

The screenshot shows the 'Verallgemeinerte lineare Modelle' dialog box with the 'Schätzung' tab selected. The dialog is organized into several sections:

- Parameterschätzung:**
 - Methode:** Hybrid (dropdown menu)
 - Maximalzahl der Iterationen bei der Fisher-Bewertung:** 1 (text input)
 - Skalenparametermethode:** Fester Wert (dropdown menu)
 - Wert:** 1 (text input)
 - Kovarianzmatrix:**
 - Modell-basierter Schätzer
 - Robuster Schätzer
 - Anfangswerte für Parameterschätzer aus einem Datenblatt abrufen
 - [Anfangswerte...](#) (button)
- Iterationen:**
 - Maximalzahl der Iterationen:** 100 (text input)
 - Maximalzahl für Schritt-Halbierung:** 5 (text input)
 - Prüfen auf Datenpunkttrennung
 - Beginn der Iteration:** 20 (text input)
 - Konvergenzkriterien:**
 - Minimum: 1E-006 (text input)
 - Typ: Absolut (dropdown menu)
 - Änderung bei den Parameterschätzern
 - Änderung in Log-Likelihood
 - Konvergenz der Hesse-Matrix
 - Singularitätstoleranz:** 1E-012 (dropdown menu)

Buttons at the bottom: OK, Einfügen, Zurücksetzen, Abbrechen, Hilfe.

Parameterschätzung. Mit den Steuerelementen in dieser Gruppe können Sie Schätzmethoden festlegen und Anfangswerte für die Parameterschätzer 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 der Fisher-Bewertung durchgeführt werden und dann zur Methode "Newton-Raphson" gewechselt wird. Wenn während der Phase "Fisher-Bewertung" 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, eine Binomialverteilung oder eine Multinomialverteilung 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 Parameterschätzer 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.

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

Konvergenzkriterien.

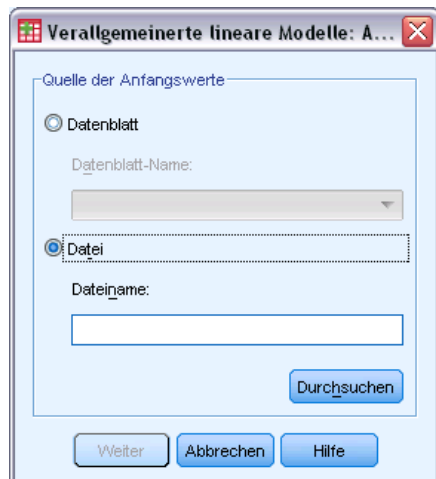
- **Parameter-Konvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzern unter dem angegebenen (positiven) Wert liegt.
- **Log-Likelihood-Konvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration 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. nichtinvertierbare) 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 ein.

Verallgemeinerte lineare Modelle – Anfangswerte

Abbildung 6-8

Verallgemeinerte lineare Modelle – Dialogfeld “Anfangswerte”



Wenn Anfangswerte angegeben wurden, müssen diese für alle Parameter (auch redundante Parameter) im Modell angegeben werden. Im Daten-Set müssen die Variablen wie folgt von links nach rechts angeordnet sein: $RowType_$, $VarName_$, $P1$, $P2$, Dabei sind $RowType_$ und $VarName_$ String-Variablen und $P1$, $P2$, ... 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 $P1$, $P2$, ... 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 Schwellenparameter, 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” in Kraft ist, müssen die Variablen mit der bzw. den Dateiaufteilungsvariablen beginnen (in der beim Erstellen der aufgeteilten Datei angegebenen Reihenfolge). Anschließend folgt $RowType_$, $VarName_$, $P1$, $P2$, ..., wie oben angegeben. Aufteilungen müssen im angegebenen Daten-Set in derselben Reihenfolge erfolgen wie im ursprünglichen Daten-Set.

Anmerkung: Die Variablennamen $P1$, $P2$, ... 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

Abbildung 6-9
Generalized Linear Models: Registerkarte "Statistiken"

The screenshot shows the 'Statistiken' (Statistics) tab of the 'Verallgemeinerte lineare Modelle' dialog box. The interface is organized into several sections:

- Modelleffekte:**
 - Analysotyp:** Typ III (dropdown menu)
 - Konfidenzniveau (%):** 95 (text input)
- Chi-Quadrat-Statistik:**
 - Wald
 - Likelihood-Quotient
- Konfidenzintervalltyp:**
 - Wald
 - Profil-Likelihood
 - Toleranzniveau:** .0001 (text input)
 - Bootstrap...** (button)
- Log-Likelihood-Funktion:** Voll (dropdown menu)
- Drucken:**
 - Zusammenfassung der Fallverarbeitung
 - Deskriptive Statistik
 - Modellinformationen
 - Statistik für Anpassungsgüte
 - Modellzusammenfassingsstatistik
 - Parameterschätzer
 - Exponentielle Parameterschätzer einbeziehen
 - Kovarianzmatrix für Parameterschätzer
 - Korrelationsmatrix für Parameterschätzer
 - Kontrastkoeffizienten-(L-)Matrizen
 - Allgemeine schätzbare Funktionen
 - Iterationsprotokoll
 - Druckintervall:** 1 (text input)
 - Lagrange-Multiplikator-Test für metrische Parameter oder negative binomiale Hilfsparameter
- Buttons:** OK, Einfügen, Zurücksetzen, Abbrechen, Hilfe

Modelleffekte.

- 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 Einflussvariablen 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 für das Konfidenzniveau einen Wert an, der über 50 und unter 100 liegt. Wald-Intervalle beruhen auf der Annahme, dass die Parameter eine asymptotische Normalverteilung aufweisen. Profile-Likelihood-Intervalle sind genauer, können jedoch sehr rechenaufwendig sein. Das Toleranzniveau für Profile-Likelihood-Intervalle ist das

Kriterium, das verwendet wird, um den zur Berechnung der Intervalle verwendeten iterativen Algorithmus anzuhalten.

- **Log-Likelihood-Funktion** Legt das Anzeigeformat der Log-Likelihood-Funktion fest. Die vollständige Funktion enthält einen zusätzlichen Term, der hinsichtlich der Parameterschätzer 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 Statistik.** Zeigt eine deskriptive Statistik und Zusammenfassungsinformationen über die abhängige Variable, die Kovariaten und die Faktoren an.
- **Modellinformationen.** Zeigt den Namen des Daten-Sets, die abhängige Variable bzw. die Ereignis- und Versuchsvariablen, die Offset-Variable, die Skalengewichtsvariable, 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.
- **Parameterschätzer.** Zeigt Parameterschätzer und die entsprechenden Teststatistiken und Konfidenzintervalle an. Wahlweise können Sie zusätzlich zu den rohen, unbearbeiteten Parameterschätzern auch potenzierte Parameterschätzer anzeigen.
- **Kovarianzmatrix für Parameterschätzer.** Zeigt die Kovarianzmatrix für die geschätzten Parameter an.
- **Korrelationsmatrix für Parameterschätzer.** Zeigt die Korrelationsmatrix für die geschätzten Parameter an.
- **Kontrastkoeffizienten-(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 Kontrastkoeffizienten-(L-)Matrizen an.
- **Iterationsprotokoll.** Zeigt das Iterationsprotokoll für Parameterschätzer und Log-Likelihood an und druckt die letzte Auswertung des Gradientenvektors und der Hesse-Matrix. Die Tabelle mit dem Iterationsprotokoll zeigt Parameterschätzer für jede n -te Iteration an, beginnend mit der 0-ten Iteration (Anfangsschätzungen). Dabei ist n der Wert des Druckintervalls. Wenn das Iterationsprotokoll angefordert wird, wird die letzte Iteration stets angezeigt, unabhängig von n .
- **Lagrange-Multiplikator-Test.** Zeigt die Statistiken für den Lagrange-Multiplikator-Test an, die zur Bewertung der Gültigkeit eines Skalenparameters dienen, der mithilfe des Pearson-Chi-Quadrats berechnet wurde oder für den bei der Normalverteilung, Gamma-Verteilung, 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

Abbildung 6-10
Generalized Linear Models: Registerkarte "Geschätzte Randmittel"

The screenshot shows the 'Verallgemeinerte lineare Modelle' dialog box with the 'Geschätzte Randmittel' tab selected. The dialog is divided into several sections:

- Typ des Modells:** A dropdown menu showing 'Modell'.
- Abhängige Variable:** A dropdown menu showing 'Schätzung'.
- Einflussvariablen:** A dropdown menu showing 'Statistiken'.
- Faktoren und Wechselwirkungen:** A list of factors and interactions with checkboxes. The factors listed are 'duration', 'treatment', and 'id'. The 'duration' and 'treatment' checkboxes are checked.
- Mittelwerte anzeigen für:** A table showing the mean values for different terms and contrasts. The table has three columns: 'Term', 'Kontrast', and 'Referenzkategorie'. The data is as follows:

Term	Kontrast	Referenzkategorie
duration	None	2
treatment	None	
duration*treatment	Keine	
- Metrisch:** Two radio buttons for calculating means: 'Mittelwerte für abhängige Variable berechnen' (selected) and 'Mittelwerte für lineare Einflussvariable berechnen'.
- Korrektur für Mehrfachvergleiche:** A dropdown menu showing 'Geringste signifikante Differenz'.
- Geschätzten Gesamtmittelwert anzeigen:** A checkbox that is currently unchecked.

At the bottom of the dialog, there are buttons for 'OK', 'Einfügen', 'Zurücksetzen', 'Abbrechen', and 'Hilfe'.

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

Faktoren und Wechselwirkungen. Diese Liste enthält Faktoren, die auf der Registerkarte "Einflussvariablen" angegeben wurden, und Wechselwirkungen 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 Wechselwirkungsterm kombiniert werden.

Mittelwerte anzeigen für. Geschätzte Mittel werden für die ausgewählten Faktoren und Wechselwirkungen 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 Wechselwirkungen zwischen Faktoren.
- **Einfach.** Vergleicht den Mittelwert jeder Stufe mit dem Mittelwert einer vorgegebenen Stufe. 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 Stufe (mit Ausnahme der ersten) mit dem Mittelwert der vorangehenden Stufen. Diese Kontraste werden auch als umgekehrte Helmert-Kontraste bezeichnet.
- **Helmert.** Vergleicht den Mittelwert jeder Stufe des Faktors (mit Ausnahme der letzten) mit dem Mittelwert der folgenden Stufen.
- **Wiederholt.** Vergleicht den Mittelwert jeder Stufe (außer der letzten) mit dem Mittelwert der folgenden Stufe.
- **Polynomial.** Vergleicht den linearen, quadratischen, kubischen Effekt usw. 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.

Skala. Geschätzte Randmittel können für die Antwort berechnet werden (auf der Grundlage der ursprünglichen Skala der abhängigen Variablen) oder für die lineare Einflussvariable (auf der Grundlage der durch die Verknüpfungsfunktion transformierten abhängigen Variablen).

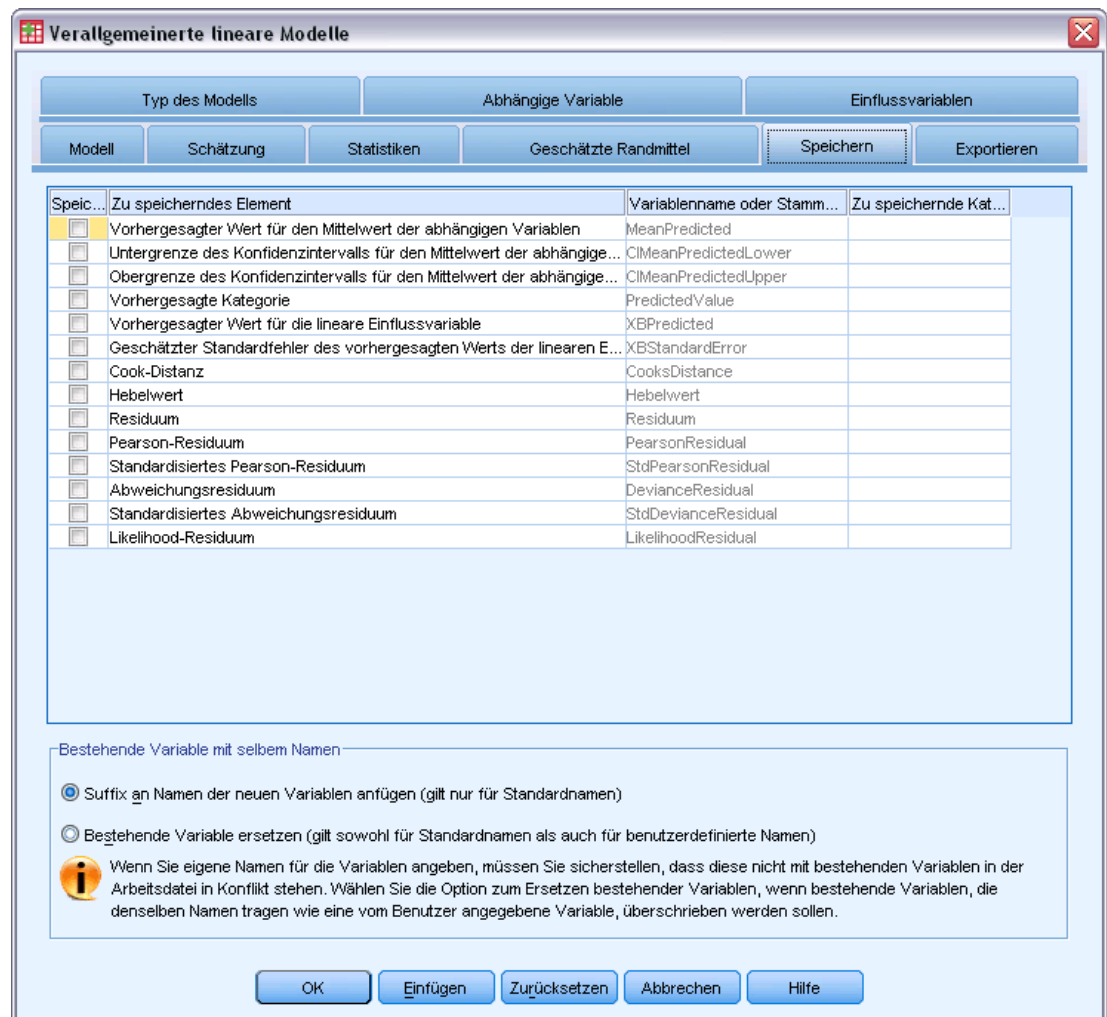
Korrektur für Mehrfachvergleiche. Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **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 sequenzielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.

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

Verallgemeinerte lineare Modelle – Speichern

Abbildung 6-11
Generalized Linear Models: Registerkarte "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 Variablenamen eindeutig machen, vermeiden.

- **Vorhergesagter Wert des Mittelwerts der Antwort.** 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 Objektbeschriftung 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 Antwort.** Speichert die Untergrenze des Konfidenzintervalls für den Mittelwert der Antwort. Bei einer multinomialen Antwortverteilung wird die Objektbeschriftung 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 Objektbeschriftung 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 die lineare Einflussvariable.** Speichert vom Modell vorhergesagte Werte für jeden Fall in der Metrik der linearen Einflussvariablen (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 Einflussvariablen.** 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 stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von 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.** Werte, die den Einfluss eines Punktes auf die Anpassung der Regression messen. Der zentrierte Wert für die Hebelwirkung bewegt sich zwischen 0 (kein Einfluss auf die Anpassung) und $(N-1)/N$.
- **Roh-Residuum.** 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 Roh-Residuums.
- **Standardisiertes Pearson-Residuum.** Das Pearson-Residuum multipliziert mit der Quadratwurzel der Inversen des Produkts des Skalenparameters und $1 - \text{Hebelwert}$ für den Fall.

- **Abweichungsresiduum.** Die Quadratwurzel des Beitrags eines Falls zur Abweichungstatistik, mit dem Vorzeichen des Roh-Residuums.
- **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 Roh-Residuums.

Verallgemeinerte lineare Modelle – Exportieren

Abbildung 6-12
Generalized Linear Models: Registerkarte "Exportieren"

Verallgemeinerte lineare Modelle

Typ des Modells Abhängige Variable Einflussvariablen

Modell Schätzung Statistiken Geschätzte Randmittel Speichern Exportieren

Modell als Daten exportieren

Ziel

Datenblatt
Name:

Datei
Durchsuchen

Als Daten exportieren

Parameterschätzer und Kovarianzmatrix
 Parameterschätzer und Korrelationsmatrix

Modell als XML exportieren
Durchsuchen

Als XML exportieren

Parameterschätzer und Kovarianzmatrix
 Nur Parameterschätzer

OK Einfügen Zurücksetzen Abbrechen Hilfe

Modell als Daten exportieren. Schreibt ein Daten-Set im IBM® SPSS® Statistics-Format, das die Parameter-Korrelations- oder -Kovarianzmatrix enthält (mit Parameterschätzern, 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 folgende Werte (und Wertelabel) an: *COV* (Kovarianzen), *CORR* (Korrelation), *EST* (Parameterschätzer), *SE* (Standardfehler), *SIG* (Signifikanzniveaus) und *DF* (Freiheitsgrade des Stichprobenplans). 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 für die Zeilentypen *COV* bzw. *CORR*, die Werte *P1*, *P2*, ... an, was einer geordneten Liste aller geschätzten Modellparameter (mit Ausnahme der Skalenparameter bzw. der negativen, binomialen Parameter) entspricht, mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter (einschließlich der Skalenparameter bzw. der negativen binomialen Parameter), mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Die Werte hängen jeweils vom Zeilentyp ab.

Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzer 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, anderenfalls 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, anderenfalls 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, Parameterschätzer, 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 ggf. die Parameter-Kovarianzmatrix (falls ausgewählt) im XML-Format (PMML). Anhand dieser Modelldatei können Sie die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Zusätzliche Funktionen beim Befehl GENLIN

Mit der Befehlssyntax können Sie auch Folgendes:

- Mit dem Unterbefehl `CRITERIA` können Sie Anfangswerte für Parameterschätzer 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 eine Untergruppe der Faktoren, für die geschätzte Randmittel angezeigt werden, mithilfe des angegebenen Kontrasttyps vergleichen lassen.

Siehe *Befehlssyntaxreferenz* für die vollständigen Syntaxinformationen.

Generalized Estimating Equations

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. Gesundheitsbehörden können mithilfe von verallgemeinerten Schätzungsgleichungen eine logistische Regression mit Messwiederholungen anpassen, um die Auswirkungen der Luftverschmutzung auf Kinder zu untersuchen.

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, Skalengewicht 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.

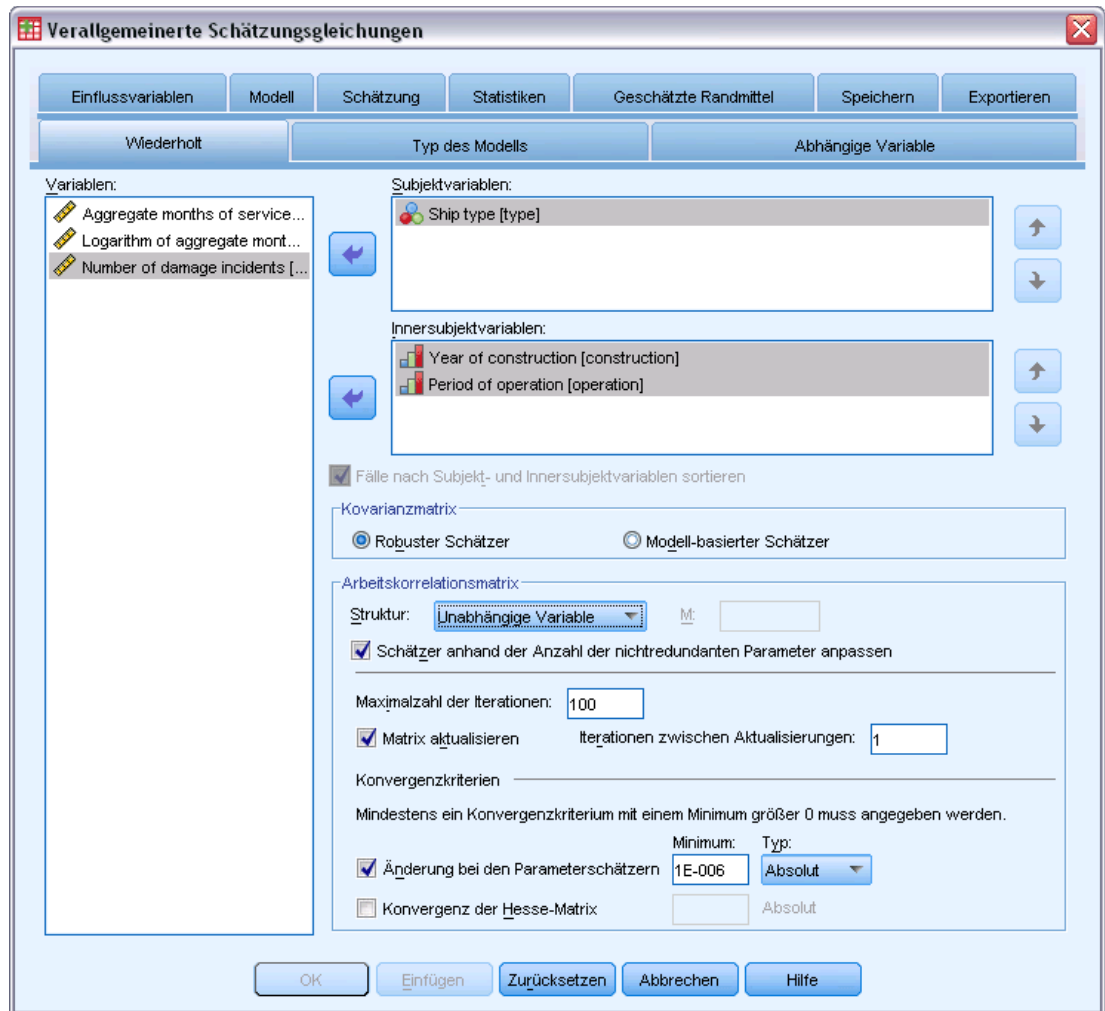
Annahmen. Es wird davon ausgegangen, dass die Fälle innerhalb der Subjekte abhängig und zwischen den verschiedenen Objekten unabhängig sind. Die Korrelationsmatrix, die die Innersubjektabhängigkeiten darstellt, wird als Teil des Modells geschätzt.

Ermitteln von verallgemeinerten Schätzungsgleichungen

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren > Generalized Linear Models > Verallgemeinerte Schätzungsgleichungen...

Abbildung 7-1
Verallgemeinerte Schätzungsgleichungen: Registerkarte "Wiederholt"



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

Die Wertekombination der angegebenen Variablen sollte die **Subjekte** innerhalb des Daten-Sets 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 Daten-Set belegen kann.

- Geben Sie auf der Registerkarte **Typ des Modells** eine Verteilungs- und Verknüpfungsfunktion an.
- Wählen Sie auf der Registerkarte **Antwort** eine abhängige Variable aus.

- ▶ Wählen Sie auf der Registerkarte **Einflussvariablen** die Faktoren und Kovariaten aus, die zur Vorhersage der abhängigen Variablen verwendet werden sollen.
- ▶ Geben Sie auf der Registerkarte **Modell** Modelleffekte mithilfe der ausgewählten Faktoren und Kovariaten 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 Daten-Set 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 zwei beliebigen Elementen ist gleich ρ für benachbarte Elemente, ρ^2 für Elemente, die durch ein drittes Element getrennt sind, usw. ρ ist beschränkt, sodass gilt: $-1 < \rho < 1$.
- **Austauschbar.** Diese Struktur weist homogene Korrelationen zwischen Elementen auf. Diese Struktur ist auch als zusammengesetzt symmetrische Struktur bekannt.
- **M-abhängig.** Aufeinanderfolgende 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. Messungen mit einem größeren Abstand gelten als

unkorreliert. Geben Sie bei der Auswahl dieser Struktur einen Wert für m an, der niedriger ist als die Ordnung der Arbeitskorrelationsmatrix.

- **Unstrukturiert (UN).** Eine ganz allgemeine Korrelationsmatrix.

Standardmäßig passt diese Prozedur die Korrelationsschätzer um die Anzahl der nichtredundanten Parameter an. Der Verzicht auf diese Anpassung kann sinnvoll sein, wenn die Schätzer gegenüber Replikationsänderungen der Daten auf der Subjektebene invariant sein sollen.

- **Maximale Anzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus für verallgemeinerte Schätzungsgleichungen vorgenommen werden. Geben Sie eine nichtnegative 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.
- **Aktualisierungsmatrix.** Die Elemente in der Arbeitskorrelationsmatrix werden anhand der Parameterschätzer geschätzt, die bei jeder Iteration des Algorithmus aktualisiert werden. Wenn die Arbeitskorrelationsmatrix überhaupt nicht aktualisiert wird, wird die ursprüngliche Arbeitskorrelationsmatrix während des gesamten Schätzungsprozesses 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.

- **Parameter-Konvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzern 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

Abbildung 7-2

Verallgemeinerte Schätzungsgleichungen: Registerkarte "Typ des Modells"

Verallgemeinerte Schätzungsgleichungen

Einflussvariablen Modell Schätzung Statistiken Geschätzte Randmittel Speichern Exportieren

Wiederholt Typ des Modells Abhängige Variable

Wählen Sie einen der unten aufgelisteten Modelltypen oder legen Sie selbst eine Kombination von Verteilungs- und Link-Funktion fest.

Metrische abhängige Variable

- Linear
- Gamma mit Log-Verknüpfungsfunktion

Ordinale abhängige Variable

- Ordinal logistisch
- Ordinal Probit

Häufigkeiten

- Poisson loglinear
- Negativ binomial mit Link-Funktion Log

Binär oder Anzahl Ereignisse

- Binär oder Anzahl Ereignisse
- Binär logistisch
- Binär Probit
- Intervallzensiertes Überleben

Mischung

- Tweedie mit Link-Funktion "Log"
- Tweedie mit Link-Funktion "Identität"

Benutzerdefiniert

- Benutzerdefiniert

Verteilung: Normal Verknüpfungsfunktion: Identität

Parameter

- Wert festlegen
- Wert: 1
- Schätzwert

Exponent:

OK Einfügen Zurücksetzen Abbrechen Hilfe

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 Antwort.

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

Ordinale Antwort.

- **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.

- **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”.

Binäre Antwort oder Ereignisse/Versuche-Daten.

- **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.

- **Tweedie mit Log-Verknüpfung.** Gibt “Tweedie” als Verteilung und “Log” als Verknüpfungsfunktion an.
- **Tweedie mit Identität-Verknüpfung.** Gibt “Tweedie” als Verteilung und “Identität” als Verknüpfungsfunktion an.

Benutzerdefiniert. Gibt 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 Daten-Set 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 nichtnegativen 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 des Hilfsparameters der negativen Binomialverteilung kann jede beliebige Zahl größer oder gleich 0 sein. Sie können einen festen Wert dafür festlegen oder den Wert durch die Prozedur schätzen lassen. Wenn der Hilfsparameter auf 0 gesetzt wird, entspricht die Verwendung dieser Verteilung der 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.** Diese Verteilung lässt sich als Anzahl der Vorkommnisse eines untersuchten Ereignisses in einem festen Zeitraum betrachten und eignet sich für Variablen mit nichtnegativen ganzzahligen Variablen mit nichtnegativen 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 Gamma-Verteilungen repräsentiert werden können; die Verteilung ist dahingehend “gemischt”, dass sie sowohl Eigenschaften von stetigen Verteilungen (nimmt nichtnegative 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 Antwort repräsentieren. Bei der abhängigen Variablen kann es sich um eine numerische Variable oder eine String-Variable handeln; sie muss mindestens zwei verschiedene gültige Datenwerte aufweisen.

Link-Funktion

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

- **Identität.** $f(x)=x$. Die abhängige Variable wird nicht transformiert. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Log-Log komplementär.** $f(x)=\log(-\log(1-x))$. Nur für die Binomialverteilung geeignet.
- **Cauchit (kumulativ).** $f(x) = \tan(\pi (x - 0,5))$ auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Nur für die Multinomialverteilung geeignet.
- **Log-Log komplementär (kumulativ).** $f(x)=\ln(-\ln(1-x))$, auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. 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.
- **Log-Log negativ (kumulativ).** $f(x)=-\ln(-\ln(x))$, auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet. Nur für die Multinomialverteilung geeignet.

- **Probit (kumulativ)**, $f(x)=\Phi^{-1}(x)$, auf die kumulative Wahrscheinlichkeit der einzelnen Kategorien der Antwort angewendet, wobei Φ^{-1} die umgekehrte kumulative Standard-Normalverteilungsfunktion ist. Nur für die Multinomialverteilung geeignet.
- **Log**, $f(x)=\log(x)$. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.
- **Log. Komplement**, $f(x)=\log(1-x)$. Nur für die Binomialverteilung geeignet.
- **Logit**, $f(x)=\log(x / (1-x))$. Nur für die Binomialverteilung geeignet.
- **Negativ binomial**, $f(x)=\log(x / (x+k^{-1}))$, wobei k der Hilfsparameter der negativen Binomialverteilung ist. Nur für die negative Binomialverteilung geeignet.
- **Log-Log negativ**, $f(x)=-\log(-\log(x))$. Nur für die Binomialverteilung geeignet.
- **Odds-Potenz**, $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$, wenn $\alpha \neq 0$. $f(x)=\log(x)$, wenn $\alpha=0$. α ist die erforderliche Zahlenangabe. Es muss sich dabei um eine reelle Zahl handeln. Nur für die Binomialverteilung geeignet.
- **Probit**, $f(x)=\Phi^{-1}(x)$, wobei Φ^{-1} die umgekehrte kumulative Standard-Normalverteilungsfunktion ist. Nur für die Binomialverteilung geeignet.
- **Exponent**, $f(x)=x^\alpha$, wenn $\alpha \neq 0$. $f(x)=\log(x)$, wenn $\alpha=0$. α ist die erforderliche Zahlenangabe. Es muss sich dabei um eine reelle Zahl handeln. Diese Verknüpfung kann mit jeder beliebigen Verteilung verwendet werden.

Verallgemeinerte Schätzungsgleichungen – Antwort

Abbildung 7-3

Verallgemeinerte Schätzungsgleichungen: Registerkarte "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 Stringvariablen oder numerische Variablen sein.
- **Anzahl der Ereignisse, die in einer Menge von Versuchen eintreten.** Wenn die Antwort eine Anzahl von Ereignissen ist, die in einer Menge von Versuchen eintreten, 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.

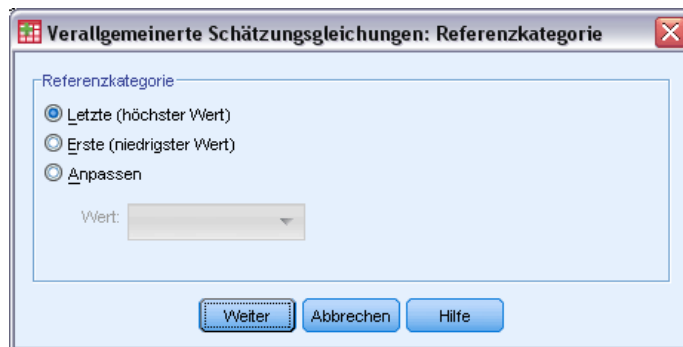
Bei den Ereignissen sollte es sich um nichtnegative Ganzzahlen und bei den Versuchen um positive Ganzzahlen handeln.

Bei ordinalen 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).

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

Verallgemeinerte Schätzungsgleichungen – Referenzkategorie

Abbildung 7-4
Verallgemeinerte Schätzungsgleichungen – Dialogfeld "Referenzkategorie"



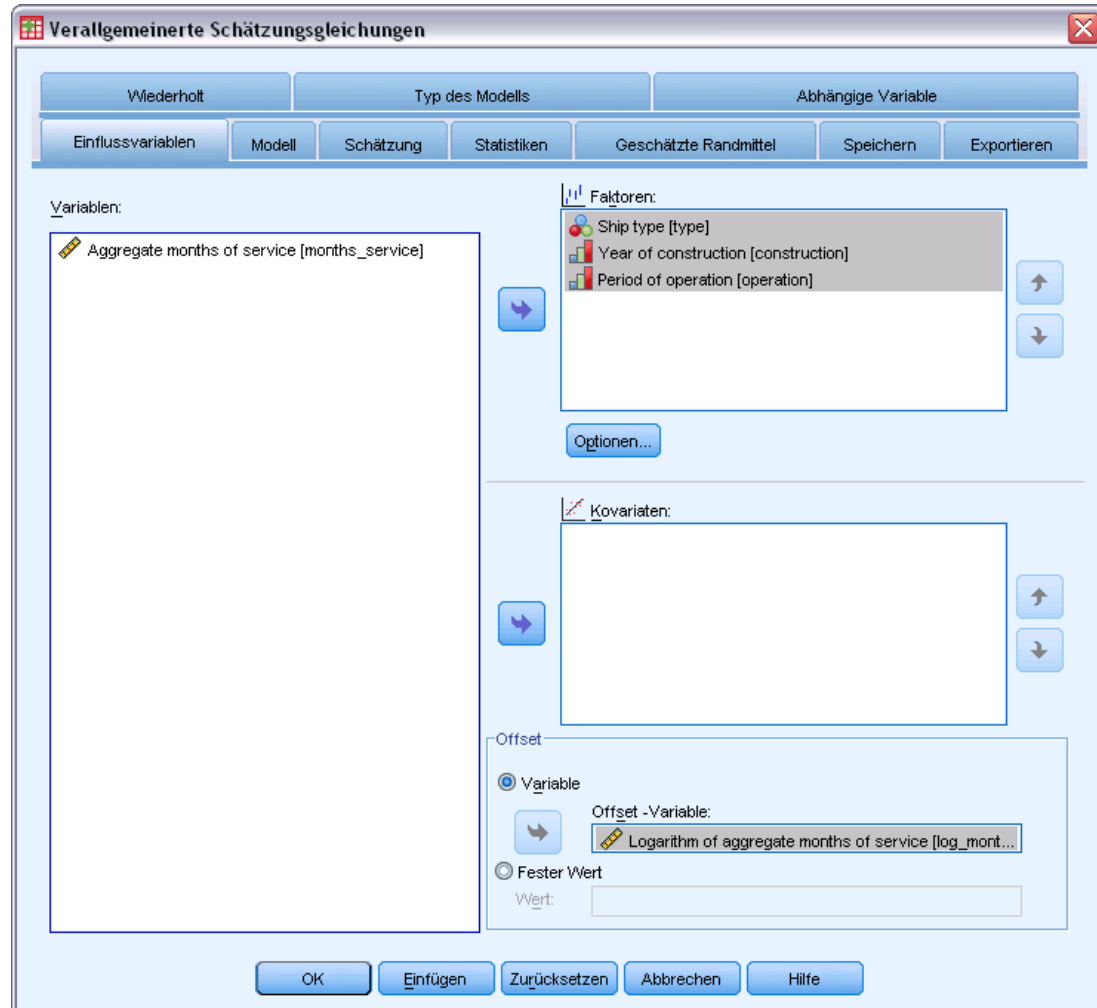
Bei binären Antworten können Sie die Referenzkategorie für die abhängige Variable auswählen. Dies kann sich auf bestimmte Ausgaben, wie beispielsweise Parameterschätzer und gespeicherte Werte, auswirken, sollte jedoch nicht die Anpassungsgüte des Modells verändern. Beispiel: Angenommen, Ihre binäre Antwort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

Verallgemeinerte Schätzungsgleichungen – Einflussvariablen

Abbildung 7-5

Verallgemeinerte Schätzungsgleichungen: Registerkarte "Einflussvariablen"



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

Faktoren. Faktoren sind kategoriale Einflussvariablen; es kann sich dabei um numerische Variablen oder Stringvariablen handeln.

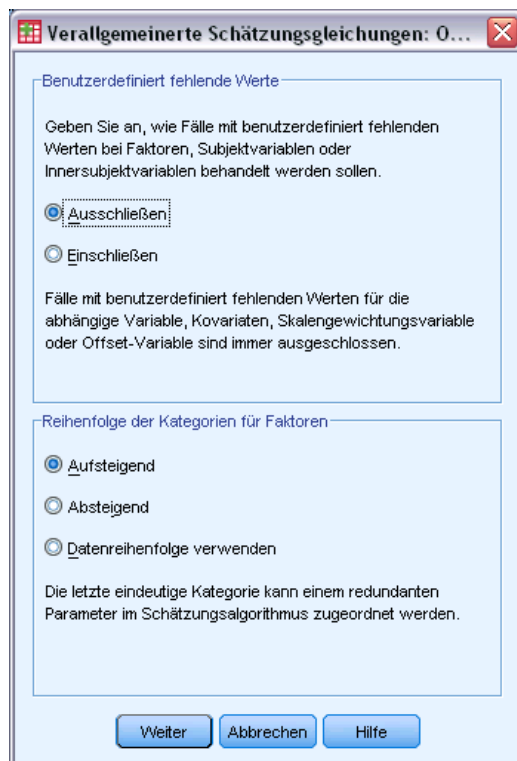
Kovariaten. Kovariaten sind metrische Einflussvariablen; sie müssen numerisch sein.

Anmerkung: 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 dieselbe Menge an Einflussvariablen während mehrerer Ausführungen der Prozedur beibehalten, um eine konsistente Anzahl an Teilgesamtheiten zu gewährleisten.

Offset. Der Term “Offset” ist eine “strukturelle” Einflussvariable. Ihr Koeffizient wird nicht vom Modell geschätzt, sondern es wird davon ausgegangen, dass er den Wert 1 aufweist. Daher werden die Werte des Offsets einfach zur linearen Einflussvariablen der abhängigen Variablen 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 poissonverteilte Response-Variable modelliert werden, wenn die Länge der Fahrpraxis als Offset-Term einbezogen wird.

Verallgemeinerte Schätzungsgleichungen – Optionen

Abbildung 7-6
Verallgemeinerte Schätzungsgleichungen – Dialogfeld “Optionen”



Diese Optionen werden auf alle auf der Registerkarte “Einflussvariablen” angegebenen Faktoren angewendet.

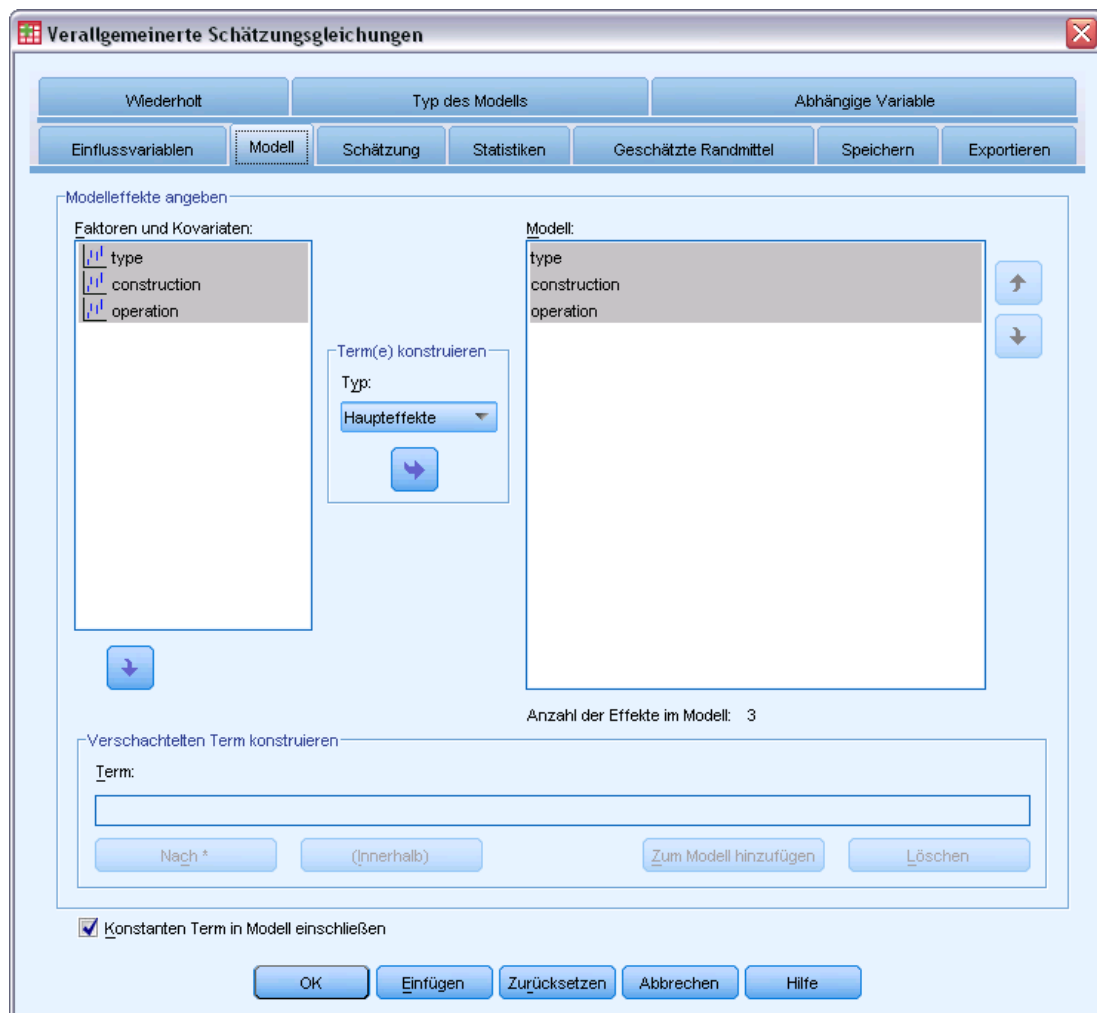
Benutzerdefinierte fehlende Werte. Faktoren müssen gültige Werte für einen Fall aufweisen, um in die Analyse aufgenommen zu werden. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Faktorvariablen als gültige Werte behandelt werden sollen.

Kategorienreihenfolge. Diese ist relevant für die Bestimmung der letzten Stufe eines Faktors, die einem redundanten Parameter im Schätzalgorithmus zugeordnet werden kann. Durch das Ändern der Kategorienreihenfolge können sich die Werte der Faktorstufen-Effekte ändern, da

diese Parameterschätzer relativ zur “letzten” 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

Abbildung 7-7
Verallgemeinerte Schätzungsgleichungen: Registerkarte “Modell”



Modell-Effekte angeben. Das Standardmodell ist “Nur konstanter Term”. Daher müssen Sie die anderen Modelleffekte explizit angeben. Wahlweise können Sie verschachtelte oder nicht verschachtelte Terme konstruieren.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

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

Mehrfaktoriell. Hiermit werden alle möglichen Wechselwirkungen und Haupteffekte der ausgewählten Variablen erzeugt.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. 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 Wechselwirkungseffekte einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Wechselwirkung 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 Kovariate 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 Schwellenparameter, die Übergangspunkte zwischen benachbarten Kategorien definieren. Die Schwellenwerte werden immer in das Modell aufgenommen.

Verallgemeinerte Schätzungsgleichungen – Schätzung

Abbildung 7-8

Verallgemeinerte Schätzungsgleichungen: Registerkarte "Schätzung"

Verallgemeinerte Schätzungsgleichungen

Wiederholt Typ des Modells Abhängige Variable

Einflussvariablen Modell **Schätzung** Statistiken Geschätzte Randmittel Speichern Exportieren

Parameterschätzung

Methode: Hybrid

Maximalzahl der Iterationen bei der Fisher-Bewertung: 1

Skalenparametermethode: Fester Wert

Wert: 1

Anfangswerte für Parameterschätzer aus einem Datenblatt abrufen

Anfangswerte...

Iterationen

Maximalzahl der Iterationen: 100

Maximalzahl für Schritt-Halbierung: 5

Prüfen auf Datenpunkttrennung

Beginn der Iteration: 20

Konvergenzkriterien

Mindestens ein Konvergenzkriterium mit einem Minimum größer 0 muss angegeben werden.

	Minimum:	Typ:
<input checked="" type="checkbox"/> Änderung bei den Parameterschätzern	1E-006	Absolut
<input type="checkbox"/> Änderung in Log-Likelihood		Absolut
<input type="checkbox"/> Konvergenz der Hesse-Matrix		Absolut

Singularitätstoleranz: 1E-012

OK Einfügen Zurücksetzen Abbrechen Hilfe

Parameterschätzung. Mit den Steuerelementen in dieser Gruppe können Sie Schätzmethoden festlegen und Anfangswerte für die Parameterschätzer 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 der Fisher-Bewertung durchgeführt werden und dann zur Methode "Newton-Raphson" gewechselt wird. Wenn während der Phase "Fisher-Bewertung" 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. Der betreffende Skalenparameterschätzer 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. Dieser Skalenparameterschätzer 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 Parameterschätzer 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 unter der Registerkarte [Wiederholt](#).

Iterationen.

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

Konvergenzkriterien.

- **Parameter-Konvergenz.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzern 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

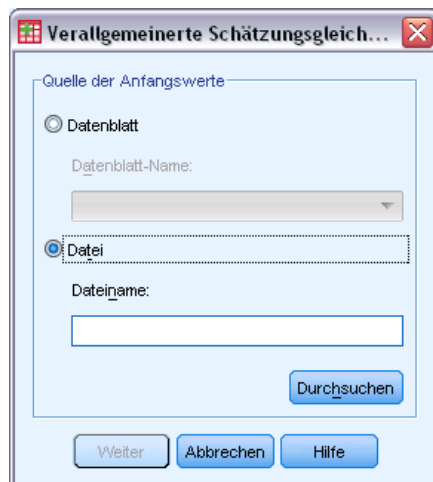
angenommen, dass eine Konvergenz vorliegt, wenn die Statistik kleiner als das Produkt aus dem angegebenen positiven Wert und dem absoluten Wert der Log-Likelihood ist.

Toleranz für Prüfung auf Singularität. Singuläre (bzw. nichtinvertierbare) 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 ein.

Verallgemeinerte Schätzungsgleichungen – Anfangswerte

Die Prozedur schätzt ein ursprüngliches generalisiertes lineares Modell. Die Schätzer aus diesem Modell werden dann als Anfangswerte für die Parameterschätzer im Bereich des linearen Modells der verallgemeinerten Schätzungsgleichungen verwendet. Anfangswerte sind für die Arbeitskorrelationsmatrix nicht erforderlich, da die Matrixelemente auf den Parameterschätzern 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.

Abbildung 7-9
Verallgemeinerte Schätzungsgleichungen – Dialogfeld “Anfangswerte”



Wenn Anfangswerte angegeben wurden, müssen diese für alle Parameter (auch redundante Parameter) im Modell angegeben werden. Im Daten-Set müssen die Variablen wie folgt von links nach rechts angeordnet sein: $RowType_$, $VarName_$, $P1$, $P2$, ... Dabei sind $RowType_$ und $VarName_$ String-Variablen und $P1$, $P2$, ... 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 $P1$, $P2$, ... 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 Schwellenparameter, 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” in Kraft ist, müssen die Variablen mit der bzw. den Dateiaufteilungsvariablen beginnen (in der beim Erstellen der aufgeteilten Datei angegebenen Reihenfolge). Anschließend folgt *RowType_*, *VarName_*, *P1*, *P2*, ..., wie oben angegeben. Aufteilungen müssen im angegebenen Daten-Set in derselben Reihenfolge erfolgen wie im ursprünglichen Daten-Set.

Anmerkung: Die Variablennamen *P1*, *P2*, ... 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

Abbildung 7-10

Verallgemeinerte Schätzungsgleichungen: Registerkarte "Statistiken"

Verallgemeinerte Schätzungsgleichungen

Wiederholt Typ des Modells Abhängige Variable

Einflussvariablen Modell Schätzung **Statistiken** Geschätzte Randmittel Speichern Exportieren

Modelleffekte

Analysetyp: Konfidenzniveau (%):

Chi-Quadrat-Statistik

Wald

Verallgemeinerter Score

Log-Quasi-Likelihood-Funktion:

Drucken

Zusammenfassung der Fallverarbeitung Kontrastkoeffizienten-(L-)Matrizen

Deskriptive Statistik Allgemeine schätzbare Funktionen

Modellinformationen Iterationsprotokoll

Statistik für Anpassungsgüte Druckintervall:

Modellzusammenfassungsverstatistik

Parameterschätzer

Exponentielle Parameterschätzer einbeziehen

Kovarianzmatrix für Parameterschätzer

Korrelationsmatrix für Parameterschätzer

Arbeitskorrelationsmatrix

OK Einfügen Zurücksetzen Abbrechen Hilfe

Modelleffekte.

- 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 Einflussvariablen im Modell zu ordnen. Typ III dagegen ist allgemeiner anwendbar. Wald-Statistiken bzw. verallgemeinerte Score-Statistiken werden anhand der Auswahl in der Gruppe "Chi-Quadrat-Statistik" berechnet.

- **Konfidenzintervalle.** Geben Sie für das Konfidenzniveau einen Wert an, der über 50 und unter 100 liegt. 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 Parameterschätzer 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 Statistik.** Zeigt eine deskriptive Statistik und Zusammenfassungsinformationen über die abhängige Variable, die Kovariaten und die Faktoren an.
- **Modellinformationen.** Zeigt den Namen des Daten-Sets, die abhängige Variable bzw. die Ereignis- und Versuchsvariablen, die Offset-Variable, die Skalengewichtsvariable, 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 Untergruppe an Einflussvariablen.
- **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.
- **Parameterschätzer.** Zeigt Parameterschätzer und die entsprechenden Teststatistiken und Konfidenzintervalle an. Wahlweise können Sie zusätzlich zu den rohen, unbearbeiteten Parameterschätzern auch potenzierte Parameterschätzer anzeigen.
- **Kovarianzmatrix für Parameterschätzer.** Zeigt die Kovarianzmatrix für die geschätzten Parameter an.
- **Korrelationsmatrix für Parameterschätzer.** Zeigt die Korrelationsmatrix für die geschätzten Parameter an.
- **Kontrastkoeffizienten-(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 Kontrastkoeffizienten-(L-)Matrizen an.
- **Iterationsprotokoll.** Zeigt das Iterationsprotokoll für Parameterschätzer und Log-Likelihood an und druckt die letzte Auswertung des Gradientenvektors und der Hesse-Matrix. Die Tabelle mit dem Iterationsprotokoll zeigt Parameterschätzer für jede n -te Iteration an, beginnend mit der 0-ten Iteration (Anfangsschätzungen). Dabei ist n der Wert des Druckintervalls. Wenn

das Iterationsprotokoll 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

Abbildung 7-11

Verallgemeinerte Schätzungsgleichungen: Registerkarte "Geschätzte Randmittel"

Verallgemeinerte Schätzungsgleichungen

Wiederholt Typ des Modells Abhängige Variable

Einflussvariablen Modell Schätzung Statistiken **Geschätzte Randmittel** Speichern Exportieren

Faktoren und Wechselwirkungen:

M	Term
<input checked="" type="checkbox"/>	type
<input checked="" type="checkbox"/>	construction
<input checked="" type="checkbox"/>	operation

Mittelwerte anzeigen für:

Term	Kontrast	Referenzkategorie
type	None	1

Metrisch

Mittelwerte für abhängige Variable berechnen

Mittelwerte für lineare Einflussvariable berechnen

Korrektur für Mehrfachvergleiche:

Geringste signifikante Differenz

Geschätzten Gesamtmittelwert anzeigen

OK Einfügen Zurücksetzen Abbrechen Hilfe

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

Faktoren und Wechselwirkungen. Diese Liste enthält Faktoren, die auf der Registerkarte “Einflussvariablen” angegeben wurden, und Wechselwirkungen 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 Wechselwirkungsterm kombiniert werden.

Mittelwerte anzeigen für. Geschätzte Mittel werden für die ausgewählten Faktoren und Wechselwirkungen 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 Wechselwirkungen zwischen Faktoren.
- **Einfach.** Vergleicht den Mittelwert jeder Stufe mit dem Mittelwert einer vorgegebenen Stufe. 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 Stufe (mit Ausnahme der ersten) mit dem Mittelwert der vorangehenden Stufen. Diese Kontraste werden auch als umgekehrte Helmert-Kontraste bezeichnet.
- **Helmert.** Vergleicht den Mittelwert jeder Stufe des Faktors (mit Ausnahme der letzten) mit dem Mittelwert der folgenden Stufen.
- **Wiederholt.** Vergleicht den Mittelwert jeder Stufe (außer der letzten) mit dem Mittelwert der folgenden Stufe.
- **Polynomial.** Vergleicht den linearen, quadratischen, kubischen Effekt usw. 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.

Skala. Geschätzte Randmittel können für die Antwort berechnet werden (auf der Grundlage der ursprünglichen Skala der abhängigen Variablen) oder für die lineare Einflussvariable (auf der Grundlage der durch die Verknüpfungsfunktion transformierten abhängigen Variablen).

Korrektur 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.

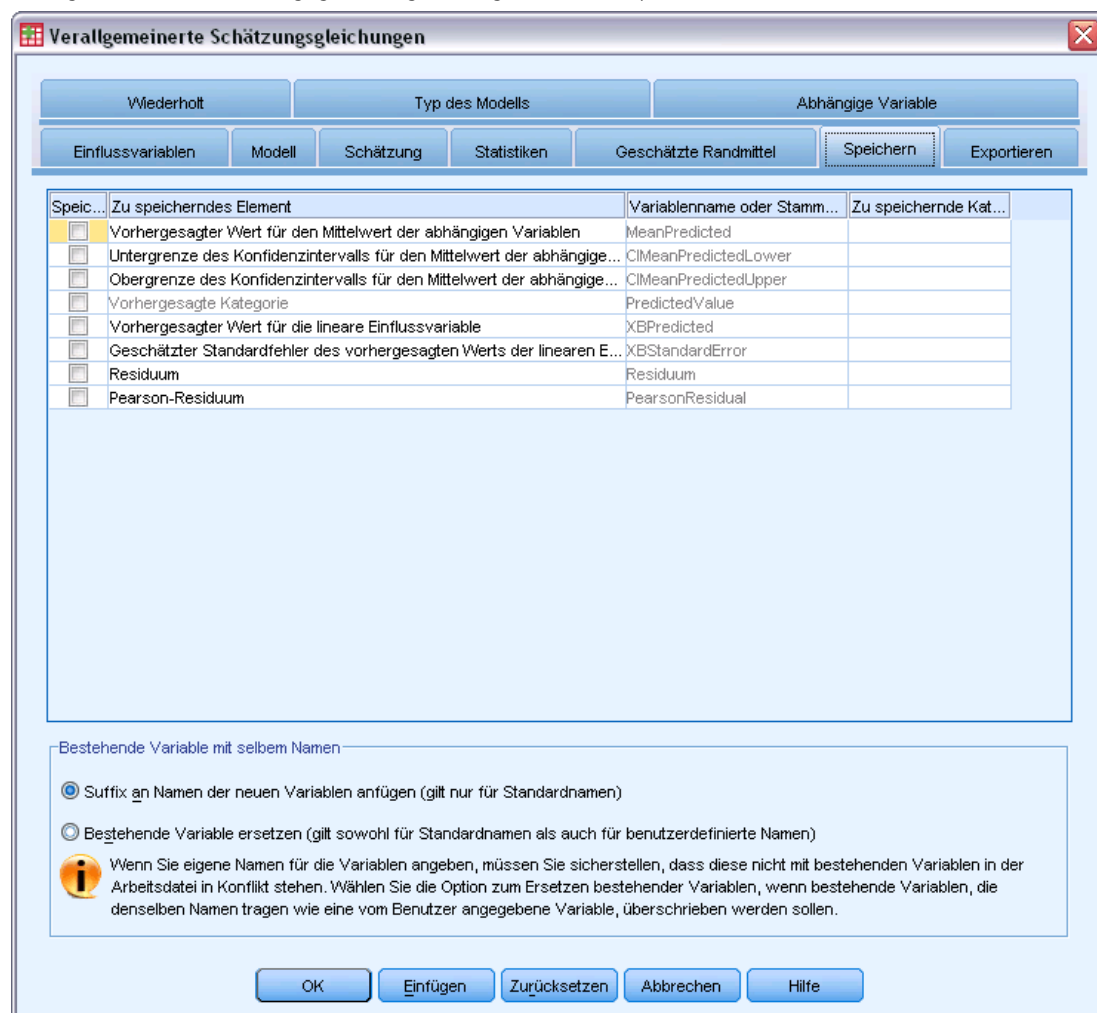
- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- **Bonferroni.** Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an,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

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

Verallgemeinerte Schätzungsgleichungen – Speichern

Abbildung 7-12

Verallgemeinerte Schätzungsgleichungen: Registerkarte "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 des Mittelwerts der Antwort.** 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 Objektbeschriftung 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 Antwort.** Speichert die Untergrenze des Konfidenzintervalls für den Mittelwert der Antwort. Bei einer multinomialen Antwortverteilung wird die Objektbeschriftung 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 Objektbeschriftung 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 die lineare Einflussvariable.** Speichert vom Modell vorhergesagte Werte für jeden Fall in der Metrik der linearen Einflussvariablen (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 Einflussvariablen.** 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.

- **Roh-Residuum.** 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 Roh-Residuums.

Verallgemeinerte Schätzungsgleichungen – Exportieren

Abbildung 7-13

Verallgemeinerte Schätzungsgleichungen: Registerkarte "Exportieren"

The screenshot shows the 'Exportieren' dialog box for Generalized Estimating Equations. The dialog has a title bar with a close button. Below the title bar are three tabs: 'Wiederholt', 'Typ des Modells', and 'Abhängige Variable'. Below these are six sub-tabs: 'Einflussvariablen', 'Modell', 'Schätzung', 'Statistiken', 'Geschätzte Randmittel', and 'Speichern'. The 'Exportieren' sub-tab is active and highlighted with a dashed border. The main area contains three sections:

- Modell als Daten exportieren:** A checkbox is checked. Below it is a 'Ziel' section with two radio buttons: 'Datenblatt' (selected) and 'Datendatei'. A 'Name:' text box is below 'Datenblatt'. A 'Durchsuchen' button is below 'Datendatei'.
- Als Daten exportieren:** Two radio buttons: 'Parameterschätzer und Kovarianzmatrix' (selected) and 'Parameterschätzer und Korrelationsmatrix'.
- Modell als XML exportieren:** A checkbox is checked. Below it is a 'Durchsuchen' button.
- Als XML exportieren:** Two radio buttons: 'Parameterschätzer und Kovarianzmatrix' (selected) and 'Nur Parameterschätzer'.

At the bottom of the dialog are five buttons: 'OK', 'Einfügen', 'Zurücksetzen', 'Abbrechen', and 'Hilfe'.

Modell als Daten exportieren. Schreibt ein Daten-Set im IBM® SPSS® Statistics-Format, das die Parameter-Korrelations- oder -Kovarianzmatrix enthält (mit Parameterschätzern, 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 folgende Werte (und Wertelabel) an: *COV* (Kovarianzen), *CORR* (Korrelation), *EST* (Parameterschätzer), *SE* (Standardfehler), *SIG* (Signifikanzniveaus) und *DF* (Freiheitsgrade des Stichprobenplans). 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 für die Zeilentypen *COV* bzw. *CORR*, die Werte *P1*, *P2*, ... an, was einer geordneten Liste aller geschätzten Modellparameter (mit Ausnahme der Skalenparameter bzw. der negativen, binomialen Parameter) entspricht, mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter (einschließlich der Skalenparameter bzw. der negativen binomialen Parameter), mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Die Werte hängen jeweils vom Zeilentyp ab.

Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzer 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, anderenfalls 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, anderenfalls 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, Parameterschätzer, 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 ggf. die Parameter-Kovarianzmatrix (falls ausgewählt) im XML-Format (PMML). Anhand dieser Modelldatei können Sie die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Zusätzliche Funktionen beim Befehl GENLIN

Mit der Befehlssyntax können Sie auch Folgendes:

- Mit dem Unterbefehl *CRITERIA* können Sie Anfangswerte für Parameterschätzer 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 eine Untergruppe der Faktoren, für die geschätzte Randmittel angezeigt werden, mithilfe des angegebenen Kontrasttyps vergleichen lassen.

Siehe *Befehlssyntaxreferenz* für die vollständigen Syntaxinformationen.

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.

Abbildung 8-1
Registerkarte "Datenstruktur"

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

Subjekte. Die Wertekombination der angegebenen kategorialen Felder sollte die Subjekte innerhalb des Daten-Sets 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 Daten-Set 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 auf der Registerkarte “Datenstruktur” 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.

Kovarianzgruppe definieren nach. Die hier festgelegten Felder definieren unabhängige Sätze von Kovarianzparametern wiederholter Effekte; einen für jede Kategorie, die durch die Kreuzklassifikation der Gruppierungsfelder definiert werden. Alle Subjekte weisen denselben Kovarianztyp auf. Subjekte innerhalb derselben Kovarianzgruppierung weisen dieselben Werte für die Parameter auf.

Kovarianztyp bei Messwiederholung. 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 (ZS)
- Diagonal
- Skalierte Identität
- Toeplitz
- Unstrukturiert (UN)
- Varianzkomponenten

Für weitere Informationen siehe Thema Kovarianzstrukturen in Anhang B auf S. 169.

Erzeugen eines verallgemeinerten linearen gemischten Modells

Für diese Funktion ist die Option “Advanced Statistics” erforderlich.

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren > Gemischte Modelle > Verallgemeinerte lineare...

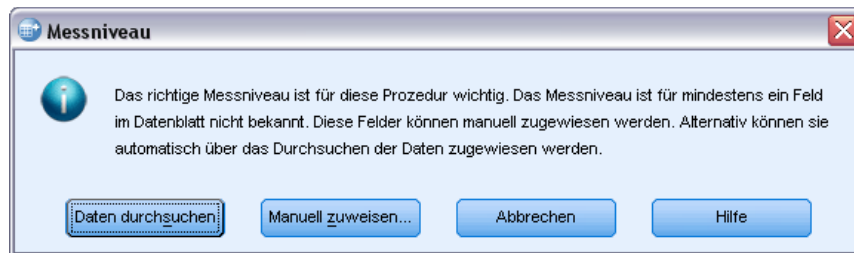
- ▶ Legen Sie die Subjektstruktur Ihres Daten-Sets auf der Registerkarte Datenstruktur fest.
- ▶ Auf der Registerkarte Felder und Effekte muss ein einzelnes Ziel vorhanden sein, das jedes beliebige Messniveau oder eine Festlegung für Ereignisse/Versuche aufweisen kann, in welchem Fall die Festlegungen 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.
- ▶ Klicken Sie auf Erstellungsoptionen, um optionale Erstellungseinstellungen anzugeben.
- ▶ Klicken Sie auf Modelloptionen, um Scores im aktiven Daten-Set zu speichern und das Modell an eine externe Datei zu exportieren.
- ▶ Klicken Sie auf Ausführen, um die Prozedur auszuführen und die Modellobjekte zu erstellen.

Felder mit unbekanntem Messniveau

Die Messniveau-Warnmeldung wird angezeigt, wenn das Messniveau für mindestens eine Variable (ein Feld) im Datenblatt unbekannt ist. Da sich das Messniveau auf die Berechnung der Ergebnisse für diese Prozedur auswirkt, müssen alle Variablen ein definiertes Messniveau aufweisen.

Abbildung 8-2

Messniveau-Warnmeldung



- **Daten durchsuchen.** Liest die Daten im aktiven Datenblatt (Arbeitsdatei) und weist allen Feldern, deren Messniveau zurzeit nicht bekannt ist, das Standardmessniveau zu. Bei großen Datenblättern kann dieser Vorgang einige Zeit in Anspruch nehmen.
- **Manuell zuweisen.** Öffnet ein Dialogfeld, in dem alle Felder mit unbekanntem Messniveau aufgeführt werden. Mit diesem Dialogfeld können Sie diesen Feldern ein Messniveau zuweisen. Außerdem können Sie in der Variablenansicht des Daten-Editors ein Messniveau zuweisen.

Da das Messniveau für diese Prozedur bedeutsam ist, können Sie erst dann auf das Dialogfeld zur Ausführung dieser Prozedur zugreifen, wenn für alle Felder ein Messniveau definiert wurde.

Ziel

Abbildung 8-3
Zieleinstellungen

Element auswählen:

Ziel

Feste Effekte

Zufällige Effekte

Gewichtung und Offset

Ziel

Vordefiniertes Ziel verwenden Benutzerdefiniertes Ziel verwenden

Ziel:

Number of convulsions

Mehr

Anzahl der Versuche als Nenner verwenden Referenzkategorie anpassen

Feld verwenden (keine) Wert verwenden

Referenzwert:

Zielverteilung und Beziehung (Verknüpfung) mit dem linearen Modell

Lineares Modell Multinomiale logistische Regression

Gamma Regression Binomiale logistische Regression

Loglinear Binär Probit

Negative binomiale Regression Intervallzensiertes Überleben

Benutzerdefiniert

Verteilung: Normal Verknüpfungsfunktion: Identität

Exponent: 1.0

Beschreibung

Beim linearen Modell wird eine Normalverteilung mit einer Identitätsverknüpfung verwendet, was nützlich ist, wenn das Ziel sich mit einer linearen Regression oder einem ANOVA-Modell vorhersagen lässt.

Mit diesen Einstellungen werden das Ziel, seine Verteilung und seine Beziehung zu den Einflussvariablen (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 jeden Datensatz sein. Bei den Ereignissen sollte es sich um nichtnegative 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 Parameterschätzer, 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 diesem Fall sollten Parameterschätzer als Bezug auf die Wahrscheinlichkeit von Kategorie 0 oder 1 *im Verhältnis* zur Wahrscheinlichkeit von Kategorie 2 interpretiert werden. Wenn Sie eine benutzerdefinierte Kategorie festlegen und Ihr Ziel über definierte Bezeichnungen verfügt, 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 kodiert war.

Zielverteilung und Beziehung (Verknüpfung) mit dem linearen Modell. Angesichts der Werte der Einflussvariablen 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 Einflussvariablen stehen. Schnellverfahren 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.
- **Gammaregression.** Gibt eine Gamma-Verteilung 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 Vorkommnissen 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 mit einer Verknüpfung vom Typ "Logit (verallgemeinert)" an, die eingesetzt werden sollte, wenn es sich bei dem Ziel um eine Antwort mit mehreren Kategorien handelt.
- **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 Daten-Set 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 oder 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 oder 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.
- **Negativ binomial.** Für die negative binominale Regression wird eine negative Binomialverteilung mit einer Log-Verknüpfung genutzt, die verwendet werden sollte, wenn das Ziel eine Anzahl an Vorkommnissen mit hoher Varianz darstellt.
- **Normalverteilung** Diese Option eignet sich für stetige Ziele, deren Werte eine symmetrische, glockenförmige Verteilung um einen Mittelwert aufweisen.
- **Poisson-Verteilung.** Diese Verteilung lässt sich als Anzahl der Vorkommnisse eines untersuchten Ereignisses in einem festen Zeitraum betrachten und eignet sich für Variablen mit nichtnegativen ganzzahligen Werten. Wenn ein Datenwert keine Ganzzahl oder kleiner als 0 ist oder fehlt, wird der entsprechende Fall nicht in der Analyse verwendet.

Verknüpfungsfunktionen

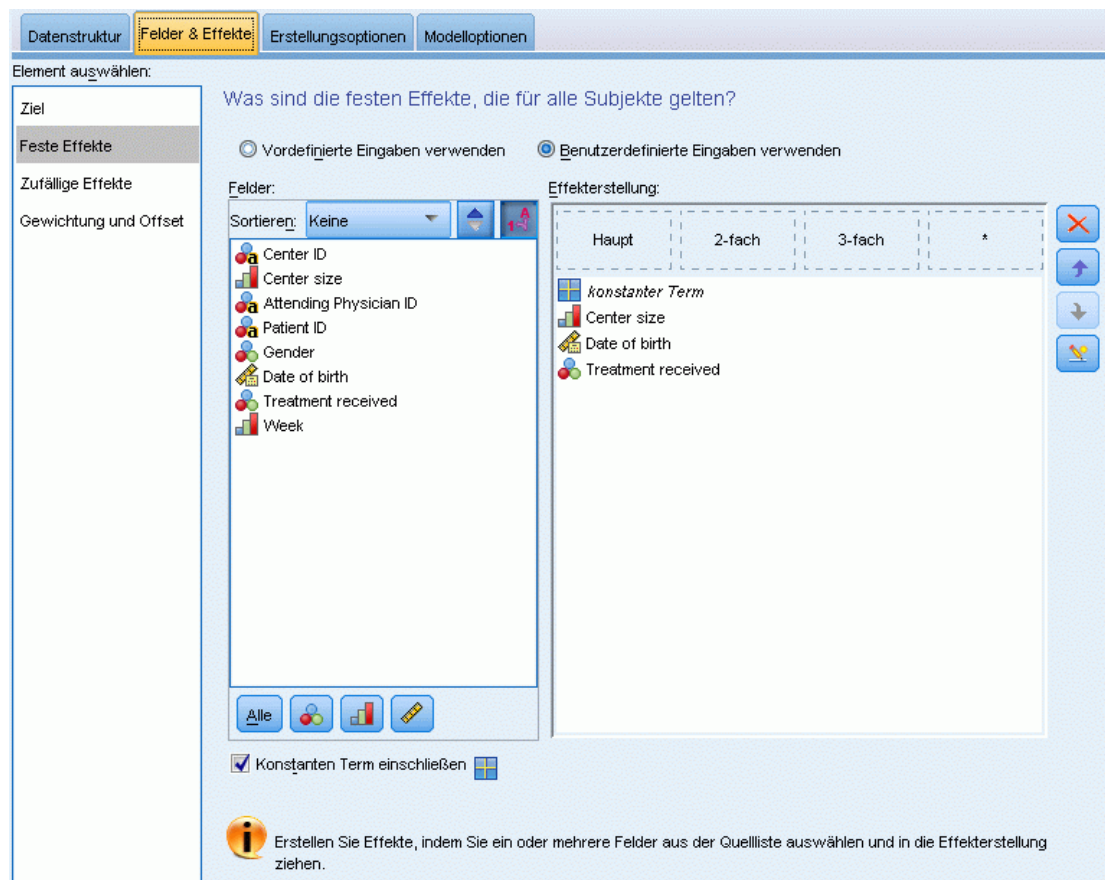
Die Verknüpfungsfunktion ist eine Transformation des Ziels, die eine Schätzung des Modells ermöglicht. Die folgenden Funktionen sind verfügbar:

- **Identität.** $f(x)=\log(x)$. Das Ziel ist nicht transformiert. Diese Verknüpfung kann abgesehen von der Multinomialverteilung mit jeder beliebigen Verteilung verwendet werden.
- **Log-Log komplementär.** $f(x)=\log(-\log(1-x))$. Nur für die Binomialverteilung geeignet.
- **Log.** $f(x)=\log(x)$. Diese Verknüpfung kann abgesehen von der Multinomialverteilung mit jeder beliebigen Verteilung verwendet werden.

- **Log. Komplement.** $f(x)=\log(1-x)$. Nur für die Binomialverteilung geeignet.
- **Logit.** $f(x)=\log(x / (1-x))$. Dies ist nur richtig bei der Binominal- oder der Multinomialverteilung; es ist die einzig gültige Verknüpfungsfunktion für die Multinomialverteilung.
- **Log-Log negativ.** $f(x)=-\log(-\log(x))$. Nur für die Binomialverteilung geeignet.
- **Probit.** $f(x)=\Phi^{-1}(x)$, wobei Φ^{-1} die umgekehrte kumulative Standard-Normalverteilungsfunktion ist. Nur für die Binomialverteilung geeignet.
- **Exponent.** $f(x)=x^\alpha$, wenn $\alpha \neq 0$. $f(x)=\log(x)$, wenn $\alpha=0$. α ist die erforderliche Zahlenangabe. Es muss sich dabei um eine reelle Zahl handeln. Diese Verknüpfung kann abgesehen von der Multinomialverteilung mit jeder beliebigen Verteilung verwendet werden.

Feste Effekte

Abbildung 8-4
Einstellungen für feste Effekte



Faktoren mit festen Effekten werden im Allgemeinen als Felder betrachtet, deren relevante Werte alle im Datenblatt (Daten-Set) dargestellt werden und zum 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 Quell-Liste 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.
- **2-fach.** Alle möglichen Paare der abgelegten Felder werden unten in der Liste der Effekte als Zweifach-Interaktionen angezeigt.
- **3-fach.** Alle möglichen Dreiergruppen der abgelegten Felder werden unten in der Liste der Effekte als Dreifach-Interaktionen angezeigt.
- *****. Die Kombination aller abgelegten Felder wird unten in der Liste der Effekte als Einzel-Interaktion angezeigt.

Die Schaltflächen rechts neben dem Effektgenerator ermöglichen Folgendes:



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,



Umsortieren von Termen im Modell mit festen Effekten durch Auswahl der Terme, die Sie umsortieren möchten, und durch Klicken auf die Pfeile nach oben bzw. unten sowie



Hinzufügen von verschachtelten Termen zum Modell mithilfe des [Benutzerdefinierten Term hinzufügen](#) -Dialogfelds durch Klicken auf die Schaltfläche "Benutzerdefinierten Term hinzufügen".

Konstanter Term einschließen. 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.

Benutzerdefinierten Term hinzufügen

Abbildung 8-5
Dialogfeld "Benutzerdefinierten Term hinzufügen"



In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. 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 Wechselwirkungseffekte, 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 Wechselwirkung 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 Kovariate ist.

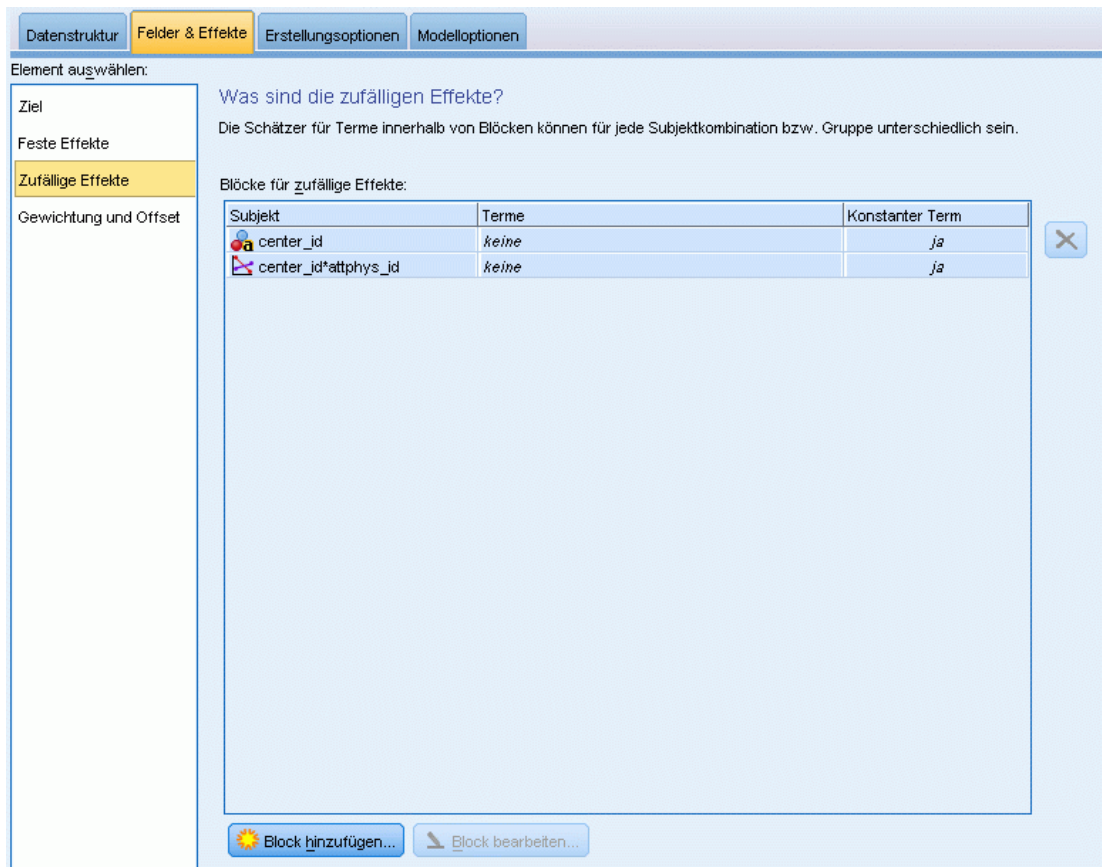
Erstellen eines verschachtelten Terms

- ▶ Wählen Sie einen Faktor oder eine Kovariate aus, der bzw. die in einem anderen Faktor verschachtelt ist, und klicken Sie auf die Pfeilschaltfläche.
- ▶ Klicken Sie auf (Innerhalb).
- ▶ Wählen Sie den Faktor aus, in dem der vorherige Faktor oder die vorherige Kovariate verschachtelt ist, und klicken Sie dann auf die Pfeilschaltfläche.
- ▶ Klicken Sie auf Term hinzufügen.

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

Zufällige Effekte

Abbildung 8-6
Einstellungen für 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:

- Zufallseffekt 1: Subjekt ist Schule (ohne Effekte, nur konstanter Term)
- Zufallseffekt 2: Subjekt ist Schule * Klasse (ohne Effekte, nur konstanter Term)

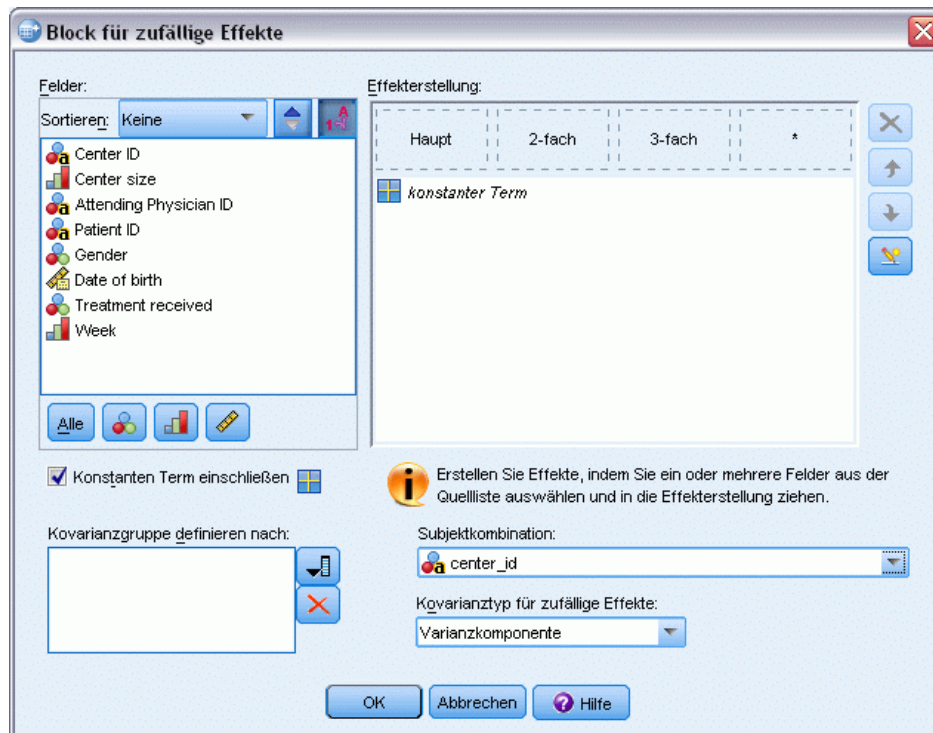
Sie können wie folgt mit Blöcken mit zufälligen Effekten arbeiten:

- ▶ Wenn Sie einen neuen Block hinzufügen möchten, klicken Sie auf **Block hinzufügen...** Dadurch wird das Dialogfeld **Block für zufällige Effekte** geöffnet.

- ▶ Wenn Sie einen bestehenden Block bearbeiten möchten, wählen Sie den gewünschten Block aus und klicken Sie auf Block bearbeiten... Dadurch wird das Dialogfeld **Block für zufällige Effekte** geöffnet.
- ▶ Um einen oder mehrere Blöcke zu löschen, wählen Sie die betreffenden Blöcke aus und klicken Sie auf die Löschschaltfläche.

Block für zufällige Effekte

Abbildung 8-7
Dialogfeld "Block für zufällige Effekte"



Geben Sie Effekte in das Modell ein, indem Sie ein oder mehrere Felder in der Quell-Liste 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. Kategoriale Felder (nominal und ordinal) werden als Faktoren im Modell verwendet und stetige Felder werden als Kovariaten verwendet.

- **Haupt.** Die abgelegten Felder werden unten in der Liste der Effekte als separate Haupteffekte angezeigt.
- **2-fach.** Alle möglichen Paare der abgelegten Felder werden unten in der Liste der Effekte als Zweifach-Interaktionen angezeigt.
- **3-fach.** Alle möglichen Dreiergruppen der abgelegten Felder werden unten in der Liste der Effekte als Dreifach-Interaktionen angezeigt.
- *****. Die Kombination aller abgelegten Felder wird unten in der Liste der Effekte als Einzel-Interaktion angezeigt.

Die Schaltflächen rechts neben dem Effektgenerator ermöglichen Folgendes:



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,



Umsortieren von Termen im Modell mit festen Effekten durch Auswahl der Terme, die Sie umsordieren möchten, und durch Klicken auf die Pfeile nach oben bzw. unten sowie



Hinzufügen von verschachtelten Termen zum Modell mithilfe des [Benutzerdefinierten Term hinzufügen](#)-Dialogfelds durch Klicken auf die Schaltfläche "Benutzerdefinierten Term hinzufügen".

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.

Kovarianzgruppe definieren nach. Die hier festgelegten Felder definieren unabhängige Sätze von Kovarianzparametern zufälliger Effekte; einen für jede Kategorie, die durch die Kreuzklassifikation der Gruppierungsfelder definiert werden. Für die einzelnen Blocks für zufällige Effekte können unterschiedliche Sätze von Gruppierungsfeldern festgelegt werden. Alle Subjekte weisen denselben Kovarianztyp auf. Subjekte innerhalb derselben Kovarianzgruppierung weisen dieselben Werte für die Parameter auf.

Subjektkombination. Damit können Sie eine Auswahl aus vorab festgelegten Kombinationen von Subjekten mit zufälligen Effekten aus der Registerkarte "Datenstruktur" treffen. Wenn beispielsweise auf der Registerkarte "Datenstruktur" 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 Zufallseffekte. 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 (ZS)
- Diagonal
- Skalierte Identität
- Toeplitz
- Unstrukturiert (UN)
- Varianzkomponenten

Für weitere Informationen siehe [Thema Kovarianzstrukturen in Anhang B auf S. 169](#).

Gewichtung und Offset

Abbildung 8-8
Einstellungen für 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 oder gleich 0 sind oder fehlen, werden in der Analyse nicht verwendet.

Offset. Der Term “Offset” ist eine “strukturelle” Einflussvariable. Ihr Koeffizient wird nicht vom Modell geschätzt, sondern es wird davon ausgegangen, dass er den Wert 1 aufweist. Daher werden die Werte des Offsets einfach zur linearen Einflussvariablen 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-Antwort modelliert werden, wenn die Fahrpraxis der Fahrer als Offset-Term eingeschlossen wird.

Erstellungsoptionen

Abbildung 8-9
Einstellungen für Erstellungsoptionen

The screenshot shows the 'Erstellungsoptionen' tab selected. The 'Sortierreihenfolge' section contains two dropdown menus, both set to 'Aufsteigend'. The 'Abbruchregeln' section has a spin box for 'Maximale Iterationen' set to 100. The 'Einstellungen nach der Schätzung' section has a 'Konfidenzniveau (%)' spin box set to 95.0. Under 'Freiheitsgrade', the first radio button 'Fest für alle Tests (Residualmethode)' is selected. Under 'Tests für feste Effekte und Koeffizienten', the first radio button 'Annehmen, dass die Modellannahmen zutreffen' is selected.

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** eine benutzerdefinierte Referenzkategorie angegeben wurde.

Abbruchregeln. Die maximale Anzahl an Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nichtnegative Ganzzahl an. Der Standardwert lautet 100.

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

- **Konfidenzniveau.** Dieses Konfidenzniveau wird zur Berechnung der Intervallschätzer der Modellkoeffizienten verwendet. Geben Sie einen Wert größer 0 und kleiner 100 ein. Der Standardwert ist 95.

- **Freiheitsgrade.** Damit wird angegeben, wie Freiheitsgrade für Signifikanztests berechnet werden. Wählen Sie die Option Fest für alle Tests (Residualmethode), wenn Ihre Stichprobe groß genug ist oder die Daten ausgeglichen sind oder das Modell einen einfacheren Kovarianztyp verwendet, z. B. “Skalierte Identität” oder “Diagonal”). Dies ist die Standardeinstellung. Wählen Sie die Option Zwischen den einzelnen Tests unterschiedlich (Satterthwaite-Approximation), wenn Ihre Stichprobe klein ist oder die Daten nicht ausgeglichen sind oder das Modell einen komplizierten Kovarianztyp verwendet, z. B. “Unstrukturiert”).
- **Test für feste Effekte und Koeffizienten.** Dies ist die Methode zur Berechnung der Kovarianzmatrix der Parameterschätzer. Wählen Sie den robusten Schätzer, wenn Sie befürchten, dass die Modellannahmen verletzt sein könnten.

Geschätzte Mittelwerte

Abbildung 8-10
Einstellungen für geschätzte Mittelwerte

Datenstruktur
Felder & Effekte
Erstellungsoptionen
Modelloptionen

Element auswählen:

Geschätzte Mittelwerte

Felder speichern

Möchten Sie das Ziel schätzen?

Geben Sie benutzerdefinierte geschätzte Mittelwerte und Kontraste an.

Terme:

Kategorial basierte Terme:	Mittelwerte schät...	Kontrasttyp	Kontrastfeld
center_size	<input checked="" type="checkbox"/>	Keines	center_size
Gender	<input type="checkbox"/>	Keines	Gender
treatment	<input type="checkbox"/>	Keines	treatment
Week	<input type="checkbox"/>	Keines	Week

Stetige Felder werden bei der Schätzung des Ziels konstant gehalten.

Felder:

Stetige Felder	Konstante	Wert
dob	Mittelwert	

Geschätzte Mittelwerte anzeigen als:

Ursprüngliche Zielskala
 Transformation einer Verknüpfungsfunktion

Anpassen für Mehrfachvergleiche mithilfe von:

Geringste signifikante Differenz

Auf dieser Registerkarte können Sie die geschätzten Randmittel für die Ebenen der Faktoren und die Wechselwirkungen zwischen Faktoren 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 Wechselwirkungen 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 die lineare Einflussvariable.

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 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 sequenziell.** Hierbei handelt es sich um ein sequenzielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Sidak (sequenziell).** Hierbei handelt es sich um ein sequenzielles schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, 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; d. h., dass bei der Methode der geringsten signifikanten Differenz mindestens so viele einzelne

Hypothesen verworfen werden wie bei der sequenziellen Sidak-Methode, bei der wiederum mindestens so viele einzelne Hypothesen verworfen werden wie bei der sequenziellen Bonferroni-Methode.

Speichern

Abbildung 8-11
Einstellungen für das Speichern

The screenshot shows the 'Modelloptionen' (Model Options) dialog box. The 'Element auswählen:' (Select element) section on the left has 'Felder speichern' (Save fields) selected. The main area is titled 'In Datenblatt speichern' (Save in data sheet) and contains several options:

- Vorhergesagte Werte (Predicted values): Feldname: PredictedValue
- Vorhergesagte Wahrscheinlichkeit für kategoriale Ziele (Predicted probability for categorical goals): Stammmame: PredictedProbability, Maximal zu speichernde Kategorien: 25
- Konfidenzintervalle (Confidence intervals): Stammmame: CI
- Pearson-Residuum (Pearson residual): Feldname: PearsonResidual
- Konfidenzen (Confidence): Feldname: Confidence, Beruht auf:
 - Wahrscheinlichkeit des vorhergesagten Werts
 - Erhöhung der Wahrscheinlichkeit gegenüber dem nächstwahrscheinlichen Wert
- Modell exportieren (Export model): Dateiname: [text box] [Durchsuchen]

At the bottom, an information icon indicates: 'Das exportierte Modell ist eine Zusammenstellung aus ein oder mehreren .xml-Dateien.'

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

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

Vorhergesagte Wahrscheinlichkeit für kategoriale Ziele. Bei kategorialen Zielen speichert dieses Schlüsselwort die vorhergesagten Wahrscheinlichkeiten der ersten n Kategorien, bis zu dem unter Maximal zu speichernde Kategorien angegebenen Wert. Der Standard-Stammmame lautet *PredictedProbability*. Um die vorhergesagte Wahrscheinlichkeit der vorhergesagten Kategorie zu speichern, speichern Sie die Konfidenz (siehe unten).

Konfidenzintervalle. Speichert die Ober- und Untergrenze des Konfidenzintervalls für den vorhergesagten Wert bzw. 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 Multinomialverteilung wird für jede Kategorie der abhängigen Variablen mit Ausnahme der letzten ([Für weitere Informationen siehe Thema Erstellungsoptionen auf S. 115.](#)) genau ein Feld erstellt. 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 die einzelnen Datensätze, das bei der Diagnose der Modellanpassung nach der Schätzung verwendet werden kann. Der Standard-Feldname lautet *PearsonResidual*.

Konfidenzen. Speichert die Konfidenz des vorhergesagten Werts 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 Standard-Feldname lautet *Konfidenz*.

Modell exportieren. Schreibt das Modell in eine externe *.zip*-Datei. Anhand dieser Modelldatei können Sie die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden. Geben Sie einen eindeutigen, gültigen Dateinamen an. Wenn die Dateispezifikation eine bestehende Datei angibt, wird diese Datei überschrieben.

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 Miniaturansichten aus.

Modellzusammenfassung

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. Außerdem werden das Akaike-Informationskriterium mit Korrektur für endliche Stichproben (AICC) und das Bayes-Informationskriterium (BIC) angezeigt.

- **Akaike (korrigiert).** Ein Maß für die Auswahl und den Vergleich von gemischten Modellen, das auf -2 (Restricted) Log-Likelihood beruht. Kleinere Werte stehen für bessere Modelle. Das AICC "korrigiert" das AIC für kleine Stichprobenumfänge. Wenn die Stichprobengröße zunimmt, konvergiert das AICC zu dem AIC.
- **Bayes.** Ein Maß für die Auswahl und den Vergleich von Modellen, das auf -2 Log-Likelihood beruht. Kleinere Werte stehen für bessere Modelle. Das BIC bestraft ebenfalls überparametrisierte Modelle, und zwar stärker als das 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 jedes Subjektfeld und Messwiederholungsfeld 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 Hitze Karte angezeigt, zuzüglich des Gesamtprozentsatzes der korrekten Werte.

Tabellenstile. Für die Diagramme sind mehrere unterschiedliche Anzeigestile verfügbar, auf die über die Dropdown-Liste Stil zugegriffen werden kann.

- **Zeilenprozentwerte.** Damit werden die Zeilenprozentsätze (die Zellenhäufigkeiten ausgedrückt als Prozentsatz der Gesamtzeilenzahl) in den Zellen angezeigt. Dies ist die Standardeinstellung.
- **Zellenhäufigkeit.** Damit werden die Zellenhäufigkeiten in den Zellen angezeigt. Die Schattierung für die Hitze Karte beruht weiterhin auf den Zeilenprozentsätzen.
- **Hitze Karte.** Damit werden keine Werte in den Zellen angezeigt, lediglich die Schattierung.
- **Komprimiert.** Damit werden keine Zeilen- oder Spaltenüberschriften oder Werte in den Zellen angezeigt. 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. Wenn mehrere kategoriale Ziele vorhanden sind, wird jedes Ziel in einer separaten Tabelle angezeigt und es gibt die Dropdown-Liste Ziel, mit der gesteuert wird, welche Ziele angezeigt werden.

Große Tabellen. Wenn das angezeigte Ziel mehr als 100 Kategorien enthält, 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 unter “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 die Standardeinstellung.
- **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 eingestellt werden, welche Effekte in der Ansicht gezeigt 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, so dass 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. Für die Diagramme 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 sind die Effekte von oben nach unten in der Reihenfolge sortiert, in der sie in den Einstellungen unter “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 unter “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. Damit werden exponentielle Koeffizienzschatzer und Konfidenzintervalle für bestimmte Modelltypen angezeigt, darunter “Binäre logistische Regression” (Binomialverteilung und Verknüpfung vom Typ “Logit”), “Nominale logistische Regression” (Multinomialverteilung und Verknüpfung vom Typ “Logit”), “Negative binomiale Regression” (Negative Binomialverteilung und Log-Verknüpfung) sowie “Loglineares Modell” (Poisson-Verteilung und Log-Verknüpfung).

Signifikanz. Mit dem Schieberegler “Signifikanz” kann eingestellt werden, welche Koeffizienten in der Ansicht gezeigt 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, so dass keine Koeffizienten basierend auf der Signifikanz herausgefiltert werden.

Kovarianzen der Zufallseffekte

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

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

- **Kovarianzwerte.** Dies ist eine Hitzekarte der Kovarianzmatrix, in der die Effekte von oben nach unten in der Reihenfolge sortiert sind, in der sie in den Einstellungen unter “Feste Effekte” angegeben wurden. Die Farben im Korrelogramm entsprechen den im Schlüssel angezeigten Zellenwerten. Dies ist die Standardeinstellung.
- **Korrelogramm.** Eine Hitzekarte der Kovarianzmatrix.
- **Komprimiert.** Eine Hitzekarte 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”.

Kovarianz-Parameter

Diese Ansicht zeigt die Kovarianzparameterschätzer 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 Schnellreferenz für die Anzahl der Parameter in den Kovarianzmatrizen für das Residuum (**R**) und den Zufallseffekt (**G**), den Rang (Anzahl der Spalten) in den Design-Matrizen für festen Effekt (**X**) und Zufallseffekt (**Z**) und die Anzahl an Subjekten, die durch die Subjektfelder für die Definition der Datenstruktur festgelegt sind.

Tabelle für Kovarianzparameter. Für den ausgewählten Effekt werden Schätzer, 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 bzw. 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 Dreifach-Interaktionen, dann die Zweifach-Interaktionen 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 Dreifach-Interaktion wird ein separates Diagramm erstellt. Alle anderen Einflussvariablen (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. Damit werden die oberen und unteren Konfidenzgrenzen für die Randmittel angezeigt, 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 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 Dreifach-Interaktion wird ein separates Diagramm erstellt. Alle anderen Einflussvariablen (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 Abstandsnetzwerkdiagramm, also um eine grafische Darstellung der Vergleichstabelle, in der die Abstände zwischen Knoten im Netzwerk den Unterschieden zwischen Stichproben entsprechen. Gelbe Linien entsprechen statistisch signifikanten Unterschieden, schwarze Linien nichtsignifikanten Unterschieden. Wenn Sie die Maus über eine Linie im Netzwerk 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. Alle anderen Einflussvariablen (Prädiktoren) werden konstant gehalten.

Wenn Kontraste angefordert wurden, wird eine weitere Tabelle mit Schätzer, Standardfehler, Signifikanztest und Konfidenzintervall für jeden Kontrast angezeigt. Bei Interaktionen wird ein separater Zeilensatz 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. Damit wird die Anzeige der oberen und unteren Konfidenzgrenzen für die Randmittel umgeschaltet, wobei das im Rahmen der Erstellungsoptionen angegebene Konfidenzniveau verwendet wird.

Layout. Damit wird das Layout für das Diagramm der paarweisen Kontraste umgeschaltet. Das Kreis-Layout zeigt weniger Kontraste als das Netz-Layout, vermeidet jedoch das Überlappen von Linien.

Modell für loglineare Analyse auswählen

Mit der Prozedur “Modellauswahl für loglineare Analyse” werden Mehr-Weg-Kreuztabellen (Kontingenztafeln) 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 die erzwungene Eingabe sowie Rückwärtselimination zur Verfügung. Bei gesättigten Modellen können Sie Parameterschätzer 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.

Statistiken. Häufigkeiten, Residuen, Parameterschätzer, Standardfehler, Konfidenzintervalle, Tests auf partiellen Zusammenhang. Bei benutzerdefinierten Modellen stehen Residuen- und Normalverteilungsdiagramme zur Verfügung.

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

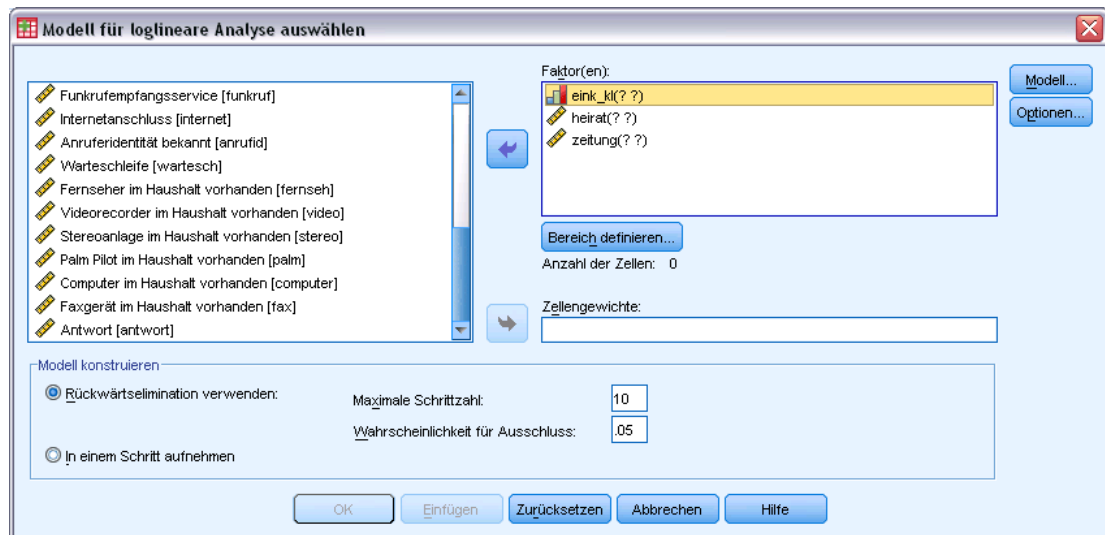
Vermeiden Sie, zahlreiche Variablen mit einer Vielzahl von Ausprägungen zu definieren. Solche Angaben können zu zahlreichen Zellen mit wenigen Beobachtungen führen, sodass die resultierenden Chi-Quadrat-Werte unbrauchbar werden.

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 umkodieren” verwenden, um String-Variablen umzukodieren. Wenn eine numerische Variable über leere Kategorien verfügt, verwenden Sie “Umkodieren”, um fortlaufende ganzzahlige Werte zu erhalten.

Anfordern einer Modellauswahl für loglineare Analyse

Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Loglinear > Modellauswahl...

Abbildung 9-1
Dialogfeld "Modell für loglineare Analyse auswählen"

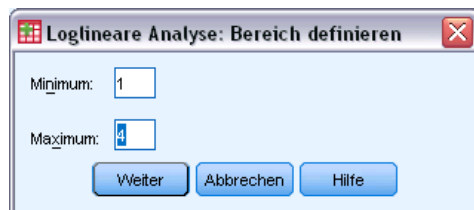


- ▶ Wählen Sie mindestens zwei numerische kategoriale Faktoren aus.
- ▶ Wählen Sie aus der Faktorenliste mindestens eine Faktorvariable aus und klicken Sie auf Bereich definieren.
- ▶ Definieren Sie den Wertebereich für jede Faktorvariable.
- ▶ Wählen Sie im Gruppenfeld "Modell konstruieren" eine Option aus.

Wahlweise können Sie eine Zellengewichtungsvariable auswählen, um Struktur-Nullen anzugeben.

Loglineare Analyse: Bereich definieren

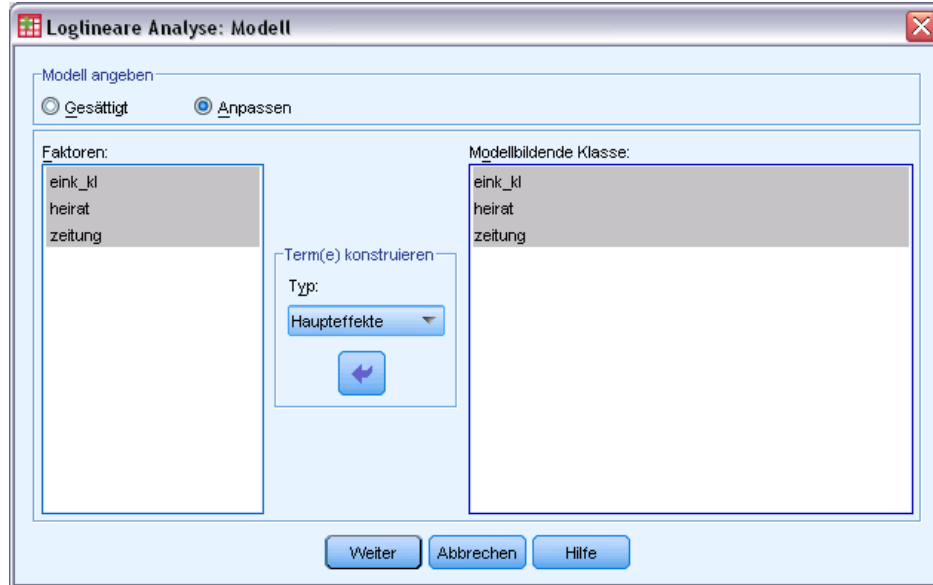
Abbildung 9-2
Dialogfeld "Loglineare Analyse: Bereich definieren"



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

Abbildung 9-3
Dialogfeld "Loglineare Analyse: Modell"



Modell angeben. Ein gesättigtes Modell enthält alle Haupteffekte der Faktoren sowie alle Wechselwirkungen 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) konstruieren" die Option Wechselwirkung aus. Dann enthält das resultierende Modell den angegebenen 3-Weg-Wechselwirkungseffekt $A*B*C$, die 2-Weg-Wechselwirkungseffekte $A*B$, $A*C$ und $B*C$ sowie die Haupteffekte für A , B und C . Definieren Sie keine verwandten Terme niedrigerer Ordnung in der modellbildenden Klasse.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

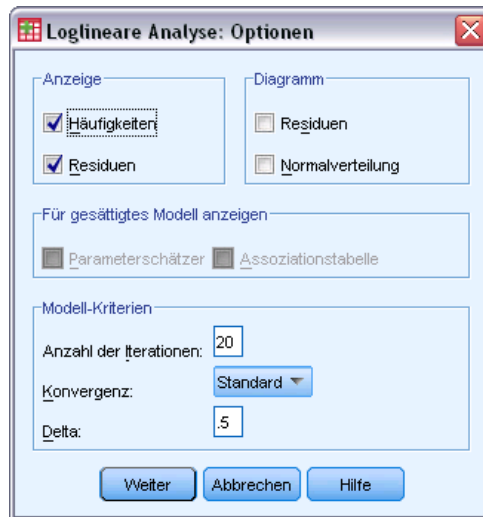
Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Modellauswahl für loglineare Analyse: Optionen

Abbildung 9-4
Dialogfeld "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 Parameterschätzer auswählen. Mithilfe von Parameterschätzern 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 rechenaufwendig.

Modell-Kriterien. Zur Anforderung von Parameterschätzern 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

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

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

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

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 Struktur-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-Quotenverhältnisse (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 in der Arbeitsdatei 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 Quotenverhältnis zufolge ergibt sich ein signifikanter Hinweis auf einen Zusammenhang.

Statistiken. Beobachtete und erwartete Häufigkeiten; unbearbeitete, korrigierte und Abweichungsresiduen; Design-Matrix; Parameterschätzer; Quotenverhältnis; Log-Quotenverhältnis; GLOR; Wald-Statistik; Konfidenzintervalle. Diagramme: korrigierte Residuen, Abweichungsresiduen und Normalverteilungsdiagramm.

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-Quotenverhältnisse. 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 Zellen Struktur-Nullen sind, nimmt die Zellenstrukturvariable einen Wert von 0 oder 1 an. Verwenden Sie zum Gewichten von aggregierten Daten keine Zellenstrukturvariable. Wählen Sie statt dessen aus dem Menü “Daten” den Befehl Fälle gewichten aus.

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

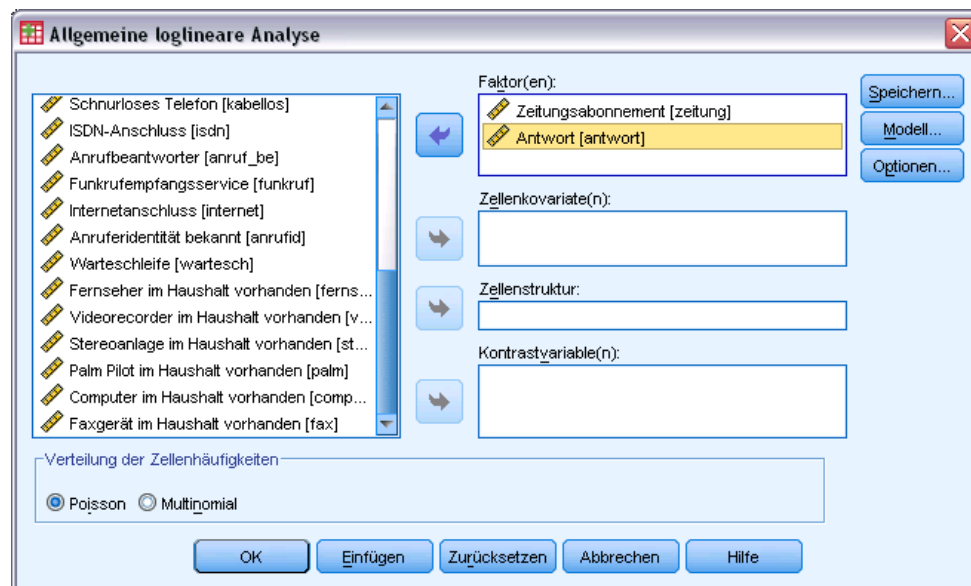
- Die gesamte Stichprobengröße wird festgesetzt oder die Analyse ist für die gesamte Stichprobengröße bedingt.
- Die Zellenhäufigkeiten sind nicht statistisch unabhängig.

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

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Loglinear > Allgemein...

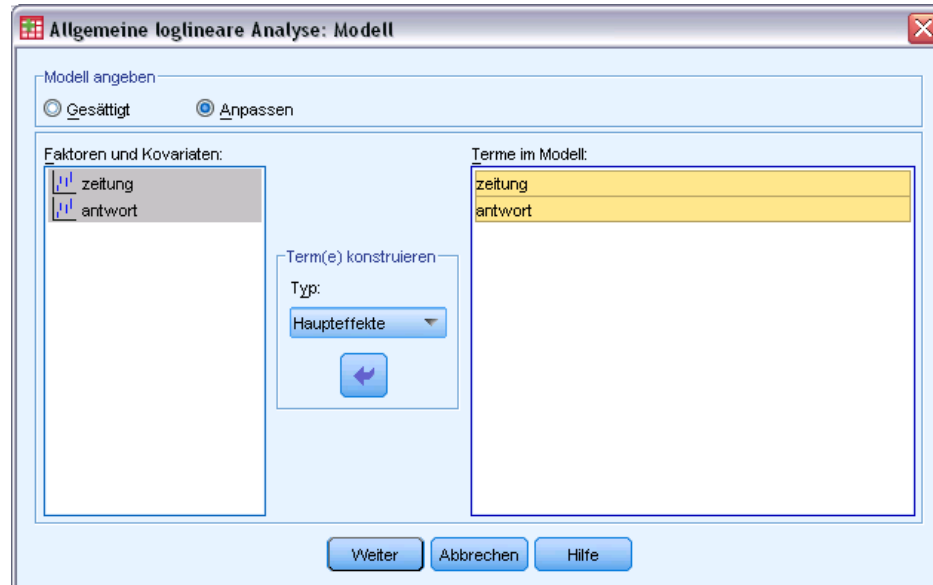
Abbildung 10-1
Dialogfeld “Allgemeine loglineare Analyse”



- ▶ Wählen Sie im Dialogfeld “Allgemeine loglineare Analyse” bis zu 10 Faktorvariablen aus. Die folgenden Optionen sind verfügbar:
 - Auswählen von Zellenkovariaten.
 - Auswählen einer Zellenstrukturvariablen, um Struktur-Nullen zu definieren, oder einen Offset-Term einzubeziehen.
 - Auswählen einer Kontrastvariablen.

Allgemeine loglineare Analyse: Modell

Abbildung 10-2
Dialogfeld "Allgemeine loglineare Analyse: Modell"



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

Faktoren und 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 Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

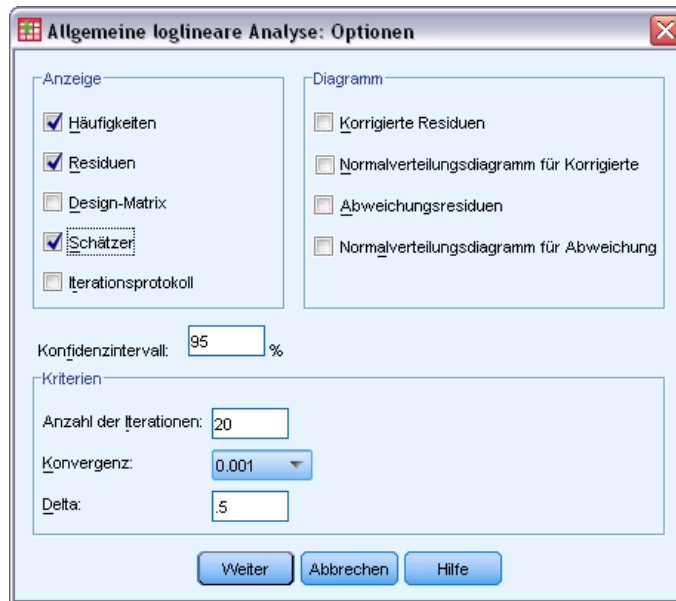
Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Allgemeine loglineare Analyse: Optionen

Abbildung 10-3
Dialogfeld "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 Design-Matrix des Modells sowie Parameterschätzer für das Modell.

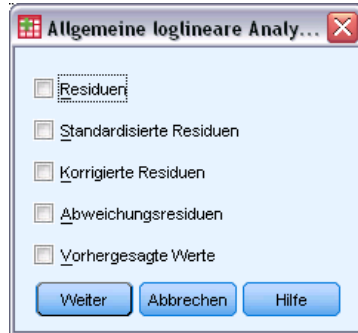
Diagramm. Als Diagramme (nur für benutzerdefinierte Modelle) sind unter anderem zwei Streudiagramm-Matrizen verfügbar (korrigierte Residuen oder Abweichungsresiduen über beobachtete und erwartete Zellenhäufigkeiten). Sie können sich auch Normalverteilungsdiagramme und zentrierte Normalverteilungsdiagramme von korrigierten Residuen oder Abweichungsresiduen anzeigen lassen.

Konfidenzintervall. Das Konfidenzintervall für die Parameterschätzer kann korrigiert werden.

Kriterien. Die Newton-Raphson-Methode wird für Maximum-Likelihood-Parameterschätzer 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

Abbildung 10-4
Dialogfeld "Allgemeine loglineare Analyse: Speichern"



Wählen Sie die Werte aus, die Sie als neue Variablen in der Arbeitsdatei 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 Kontingenztafel), auch wenn die Daten im Daten-Editor nach einzelnen Beobachtungen aufgezeichnet sind. Wenn Sie Residuen oder Vorhersagewerte für nicht aggregierte Daten speichern, wird der gespeicherte Wert einer Zelle in der Kontingenztafel für jeden in dieser Zelle vorhandenen Fall in den Daten-Editor 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, korrigierte und Abweichungsresiduen. Außerdem können die vorhergesagten Werte gespeichert werden.

- **Residuen.** Auch bekannt als einfaches Residuum oder Roh-Residuum. Die Differenz zwischen der beobachteten und der erwarteten Häufigkeit in der Zelle.
- **Standardisierte Residuen.** Der Quotient aus dem Residuum und einem Schätzer seines Standardfehlers. Standardisierte Residuen werden auch als Pearson-Residuen bezeichnet.
- **Korrigierte Residuen.** Der Quotient aus dem standardisierten Residuum und seinem geschätzten Standardfehler. Wenn das ausgewählte Modell zutrifft, sind die korrigierten Residuen asymptotisch standardnormalverteilt und daher für den Test auf Normalverteilung den standardisierten Residuen vorzuziehen.
- **Abweichungsresiduen.** Die mit Vorzeichen versehene Quadratwurzel eines einzelnen Beitrags zum Likelihood-Quotienten-Chi-Quadrat (G-Quadrat), wobei das Vorzeichen das Vorzeichen des Residuums ist (beobachtete Anzahl minus erwartete Anzahl). Abweichungsresiduen sind asymptotisch standardnormalverteilt.

Zusätzliche Funktionen beim Befehl GENLOG

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- 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 Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

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, also Zellenkovariaten, können stetig sein, werden aber nicht auf einer fallweisen Basis zugewiesen. Das gewichtete Kovariatenmittel einer Zelle wird dieser Zelle zugewiesen. Der Logarithmus der Quote 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 Struktur-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-Quotenverhältnisse (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 in der Arbeitsdatei 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 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.

Statistiken. Beobachtete und erwartete Häufigkeiten; unbearbeitete, korrigierte und abweichende Residuen; Design-Matrix; Parameterschätzer; verallgemeinertes Log-Quotenverhältnis; Wald-Statistik; Konfidenzintervalle. Diagramme: korrigierte Residuen, abweichende Residuen und Normalverteilungsdiagramme.

Daten. Die abhängigen Variablen sind kategorial. Faktoren sind kategorial. Zellenkovariaten können stetig sein, aber wenn eine Kovariate im Modell enthalten ist, wird der mittleren Kovariatenwert für Fälle einer Zelle dieser Zelle zugewiesen. Kontrastvariablen sind stetig. Sie werden verwendet, um das verallgemeinerte Log-Quotenverhältnis (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 Struktur-Nullen darstellen, weist die Zellenstrukturvariable einen Wert von entweder 0 oder 1 auf. Verwenden Sie keine Zellenstrukturvariable, um aggregierte Daten zu gewichten. Verwenden Sie statt dessen 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:

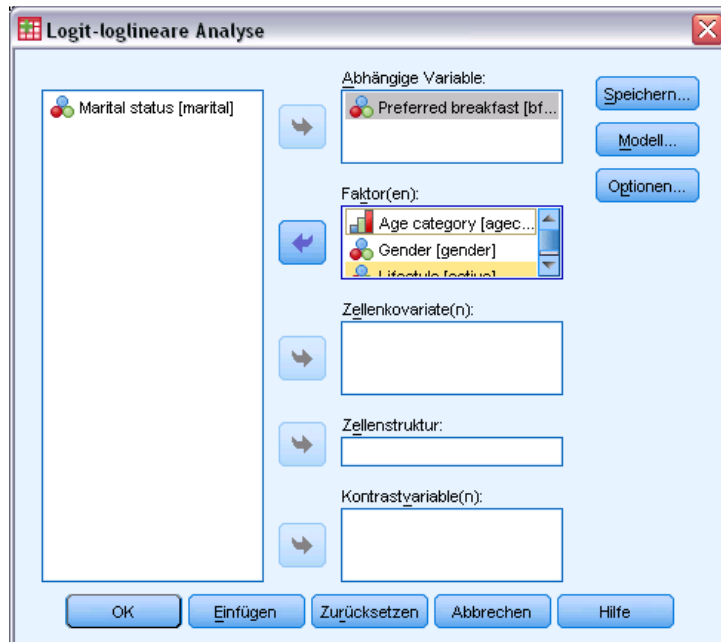
- Die gesamte Stichprobengröße wird festgesetzt oder die Analyse ist für die gesamte Stichprobengröße bedingt.
- Die Zellenhäufigkeiten sind nicht statistisch unabhängig.

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

Berechnen einer Logit-loglinearen Analyse

- Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Loglinear > Logit...

Abbildung 11-1
Dialogfeld "Logit-loglineare Analyse"



- Wählen Sie im Dialogfeld "Logit-loglineare Analyse" mindestens eine abhängige Variable aus.
 - Wählen Sie mindestens eine Faktorvariable aus.
- Die Gesamtanzahl der abhängigen Variablen und Faktorvariablen muss kleiner oder gleich 10 sein.

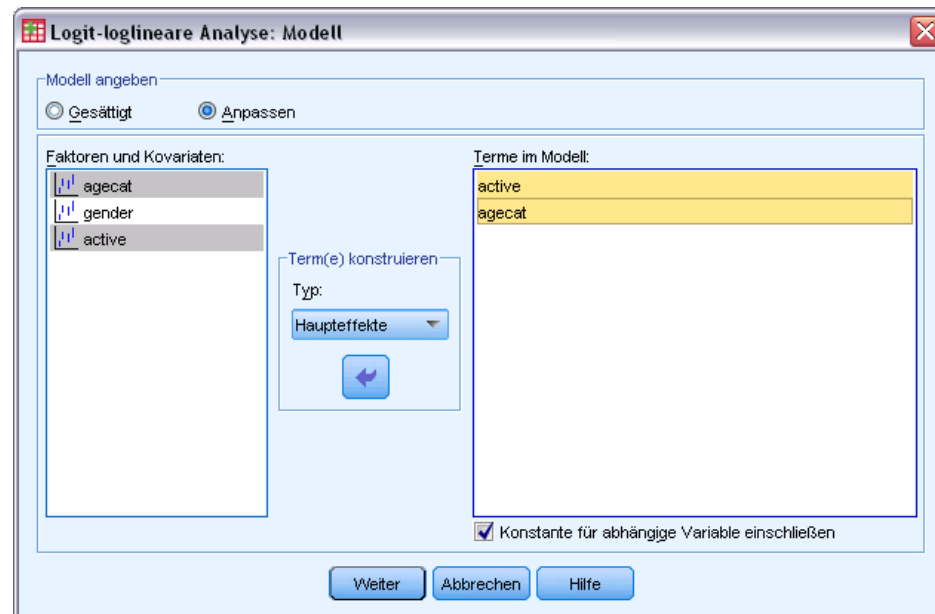
Die folgenden Optionen sind verfügbar:

- Auswählen von Zellenkovariaten.
- Auswählen einer Zellenstrukturvariablen, um Struktur-Nullen zu definieren, oder einen Offset-Term einzubeziehen.
- Auswählen einer oder mehrerer Kontrastvariablen.

Logit-loglineare Analyse: Modell

Abbildung 11-2

Dialogfeld "Logit-loglineare Analyse: Modell"



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

Faktoren und 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 Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

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 Einheitsterm (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

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

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

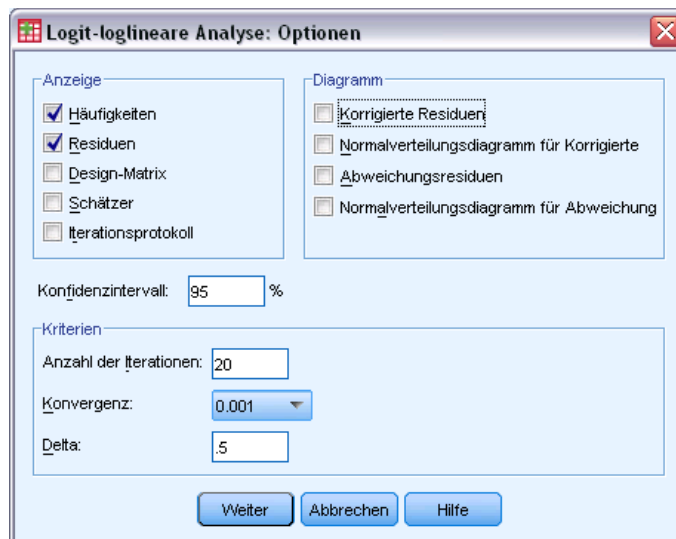
Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Logit-loglineare Analyse: Optionen

Abbildung 11-3
Dialogfeld "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. Es können verschiedene Statistiken angezeigt werden: beobachtete und erwartete Zellenhäufigkeit; unbearbeitete, korrigierte und abweichende Residuen; eine Design-Matrix des Modells und Parameterschätzer für das Modell.

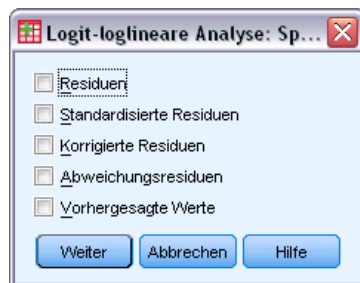
Diagramm. Für benutzerdefinierte Modelle sind zwei Streudiagramm-Matrizen verfügbar (korrigierte Residuen oder Abweichungsresiduen über beobachtete und erwartete Zellenhäufigkeiten). Sie können sich auch Normalverteilungsdiagramme und zentrierte Normalverteilungsdiagramme von korrigierten Residuen oder Abweichungsresiduen anzeigen lassen.

Konfidenzintervall. Das Konfidenzintervall für die Parameterschätzer kann korrigiert werden.

Kriterien. Die Newton-Raphson-Methode wird für Maximum-Likelihood-Parameterschätzer 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.

Logit-loglineare Analyse: Speichern

Abbildung 11-4
Dialogfeld “Logit-loglineare Analyse: Speichern”



Wählen Sie die Werte aus, die Sie als neue Variablen in der Arbeitsdatei 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 Kontingenztafel), auch wenn die Daten im Daten-Editor als einzelne Beobachtungen aufgezeichnet wurden. Wenn Sie Residuen oder Vorhersagewerte für nicht aggregierte Daten speichern, wird der gespeicherte Wert einer Zelle in der Kontingenztafel für jeden in dieser Zelle vorhandenen Fall in den Daten-Editor 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, korrigierte und Abweichungsresiduen. Außerdem können die vorhergesagten Werte gespeichert werden.

- **Residuen.** Auch bekannt als einfaches Residuum oder Roh-Residuum. Die Differenz zwischen der beobachteten und der erwarteten Häufigkeit in der Zelle.

- **Standardisierte Residuen.** Der Quotient aus dem Residuum und einem Schätzer seines Standardfehlers. Standardisierte Residuen werden auch als Pearson-Residuen bezeichnet.
- **Korrigierte Residuen.** Der Quotient aus dem standardisierten Residuum und seinem geschätzten Standardfehler. Wenn das ausgewählte Modell zutrifft, sind die korrigierten Residuen asymptotisch standardnormalverteilt und daher für den Test auf Normalverteilung den standardisierten Residuen vorzuziehen.
- **Abweichungsresiduen.** Die mit Vorzeichen versehene Quadratwurzel eines einzelnen Beitrags zum Likelihood-Quotienten-Chi-Quadrat (G-Quadrat), wobei das Vorzeichen das Vorzeichen des Residuums ist (beobachtete Anzahl minus erwartete Anzahl). Abweichungsresiduen sind asymptotisch standardnormalverteilt.

Zusätzliche Funktionen beim Befehl GENLOG

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- 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 Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Sterbetafeln (Life Tables Options)

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 schließt gewöhnlich einige Fälle ein, für die kein zweites Ereignis aufgezeichnet ist. Zum Beispiel können Arbeitnehmer der Firma beim Abschluss der Studie weiterhin in der Firma beschäftigt sein. Dies kann aus verschiedenen Gründen auftreten: In einigen Fällen tritt das Ereignis nicht vor Abschluss der Studie 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. die Beurlaubung eines krank gewordenen Beschäftigten). 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 Endereignisses 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 Nikotinplaster-Therapie besser geeignet als eine herkömmliche Nikotinplaster-Therapie, 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 Sterbetafelns 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 Hazard-Funktionen auch grafisch darstellen und diese visuell mit detaillierteren Informationen vergleichen.

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

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

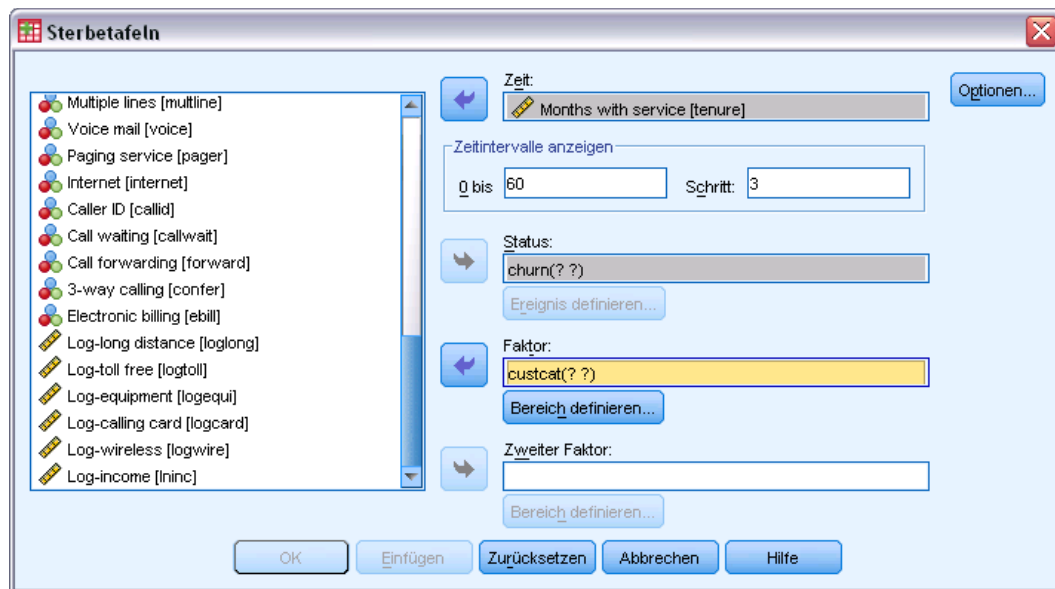
Annahmen. Wahrscheinlichkeiten für das untersuchte Ereignis dürfen nur von der Zeit nach dem Anfangsereignis abhängen. Sie werden in Bezug auf die absolute Zeit als stabil angenommen. Dies bedeutet, dass die in die Studie zu unterschiedlichen Zeiten aufgenommenen Fälle (wie beispielsweise Patienten mit 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

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Überlebenswahrscheinlichkeit > Sterbetafeln...

Abbildung 12-1
Dialogfeld "Sterbetafeln"



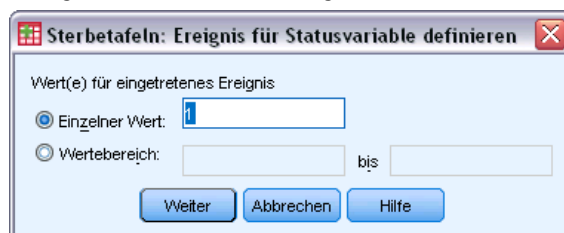
- ▶ Wählen Sie eine **numerische** Überlebensvariable aus.
- ▶ Legen Sie die zu untersuchenden Zeitintervalle fest.
- ▶ Wählen Sie eine Statusvariable aus, um die Fälle zu definieren, für die ein Endereignis stattgefunden hat.
- ▶ Klicken Sie auf Ereignis definieren, um den Wert der Statusvariablen festzulegen, 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: Ereignis für Statusvariable definieren

Abbildung 12-2
Dialogfeld "Sterbetafeln: Ereignis für Statusvariable definieren"

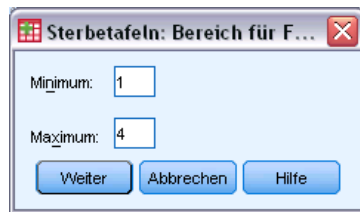


Wenn die Statusvariable den bzw. die ausgewählten Wert(e) annimmt, bedeutet dies, dass das Endereignis für diese Fälle eingetreten ist. Alle anderen Fälle werden als zu zensieren betrachtet. Geben Sie einen einzelnen Wert oder einen Bereich von Werten ein, die das zu untersuchende Ereignis kennzeichnen.

Sterbetafel(n): Bereich für Faktorvariable definieren

Abbildung 12-3

Dialogfeld "Sterbetafel(n): Bereich für Faktorvariable definieren"

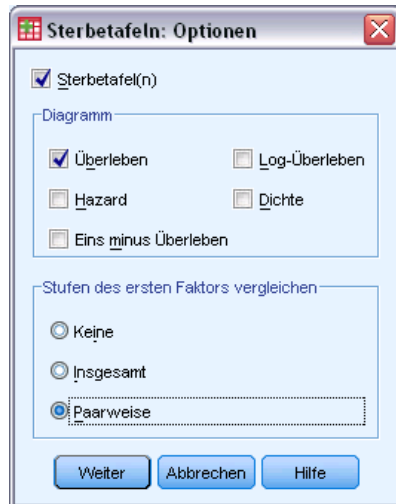


Alle Fälle, bei denen die Faktorvariable Werte im festgelegten Bereich annimmt, werden in die Analyse eingeschlossen. Für jeden eindeutigen Wert innerhalb des Bereichs werden getrennte Tabellen (und Diagramme, falls gewünscht) erzeugt.

Sterbetafel(n): Optionen

Abbildung 12-4

Dialogfeld "Sterbetafel(n): Optionen"



Sie können verschiedene Aspekte der Analyse von Sterbetafel(n) beeinflussen.

Sterbetafel(n). Um die Anzeige von Sterbetafel(n) in der Ausgabe zu unterdrücken, deaktivieren 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.

- **Überlebenswahrscheinlichkeit.** Zeigt die kumulative Überlebensfunktion auf einer linearen Skala an.
- **Log-Überleben.** Zeigt kumulative Überlebensfunktion auf einer logarithmischen Skala an.
- **Kumulative Hazardrate.** Zeigt die kumulative Hazard-Funktion auf einer linearen Skala an.
- **Dichte.** Zeigt die Dichte-Funktion an.
- **Eins minus Überleben.** Erzeugt ein Diagramm der Werte "1 - Überlebensfunktion" auf einer linearen Skala.

Stufen des ersten Faktors vergleichen. Wenn Sie über eine Kontroll-Variable 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

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

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

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

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, welche 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 untersuchende Behandlung eine Verbesserung gegenüber der traditionellen Therapie darstellt. Sie können die Überlebensverteilung oder Hazard-Funktionen auch grafisch darstellen und diese visuell mit detaillierteren Informationen vergleichen.

Statistiken. Überlebenstabelle, einschließlich Zeit, Status, kumulierte Überlebenswahrscheinlichkeit und Standardfehler, kumulierte Ereignisse sowie die verbleibende Anzahl; Mittelwert und Median der Überlebenszeit mit Standardfehler und 95%-Konfidenzintervall. Diagramme: Überlebensverteilung, Hazard, Log-Überlebensdiagramm, Eins minus Überleben.

Daten. Die Zeitvariable muss stetig sein, die Statusvariable kann kategorial oder stetig sein und die Faktor- sowie Schichtenvariablen 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 Hazard-Funktion 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

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Überlebenswahrscheinlichkeit > Kaplan-Meier...

Abbildung 13-1
Dialogfeld “Kaplan-Meier”



- ▶ Wählen Sie eine Zeitvariable aus.
- ▶ Wählen Sie eine Statusvariable zum Bestimmen von Fällen aus, für welche das Endereignis aufgetreten ist. Diese Variable kann numerisch oder ein **kurzer String** sein. Klicken Sie dann auf Ereignis definieren.

Wahlweise können Sie eine Faktorvariable auswählen, um Gruppendifferenzen zu untersuchen. Sie können auch eine Schichtenvariable auswählen, mit der sich für jedes Niveau (Schicht) der Variablen getrennte Analysen erzeugen lassen.

Kaplan-Meier: Ereignis für Statusvariable definieren

Abbildung 13-2
Dialogfeld "Kaplan-Meier: Ereignis für Statusvariable definieren"

Geben Sie den Wert oder die Werte ein, die anzeigen, dass das Endereignis 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

Abbildung 13-3
Dialogfeld "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 zu bestimmen, welche Vergleiche durchgeführt werden sollen: "Gemeinsam über Schichten", "Für jede Schicht", "Paarweise über Schichten" oder "Paarweise für jede Schicht".

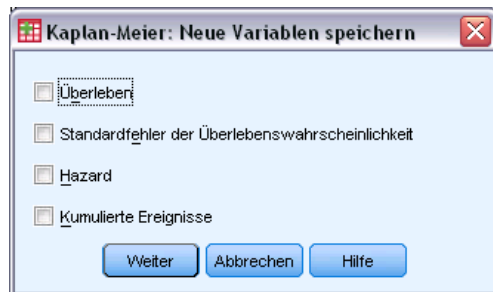
- **Log-Rang.** Ein Test, der die Gleichheit der Überlebensverteilungen vergleicht. Alle Zeitpunkte werden in diesem Test gleich gewichtet.

- **Breslow-Test.** Ein Test, der die Gleichheit der Überlebensverteilungen vergleicht. Die Zeitpunkte werden mit der Anzahl der zu den einzelnen Zeitpunkten gefährdeten Fälle gewichtet.
- **Tarone-Ware-Test.** Ein Test, der die Gleichheit der Überlebensverteilungen vergleicht. Die Zeitpunkte werden mit der Quadratwurzel der Anzahl der zu den einzelnen Zeitpunkten gefährdeten Fälle gewichtet.
- **Gemeinsam 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 Stratifikationsvariable (Schichtungsvariable) vorhanden ist, werden die Tests nicht durchgeführt.
- **Paarweise für jede Schicht.** Vergleicht jedes unterschiedliche Paar von Faktorstufen auf jeder Schicht. Paarweise Trendtests sind nicht verfügbar. Wenn keine Stratifikationsvariable (Schichtungsvariable) vorhanden ist, werden die Tests nicht durchgeführt.

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.

Kaplan-Meier: Neue Variablen speichern

Abbildung 13-4
Dialogfeld "Kaplan-Meier: Neue Variablen speichern"



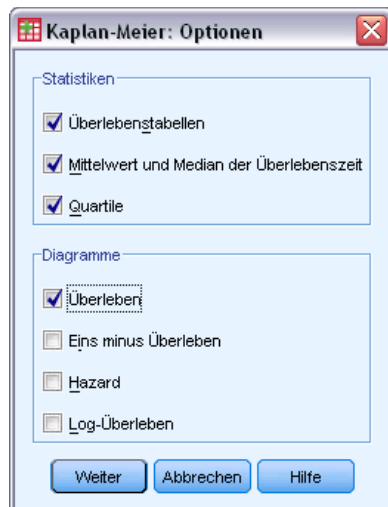
Sie können die Informationen aus der Kaplan-Meier-Tabelle als neue Variablen speichern, die dann in nachfolgenden Analysen zum Testen von Hypothesen oder Überprüfen von Annahmen verwendet werden können. Es können die Überlebensverteilung, der Standardfehler der Überlebensverteilung sowie Hazard- und kumulierte Ereignisse als neue Variablen gespeichert werden.

- **Überlebenswahrscheinlichkeit.** Schätzer der kumulierten 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 Überlebenswahrscheinlichkeit.** Standardfehler des Schätzers der kumulierten 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.
- **Kumulative Hazardrate.** Schätzer der kumulierten Hazard-Funktion. 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.
- **Kumulierte Ereignisse.** Kumulierte Ereignishäufigkeit, wenn die Fälle nach ihren Überlebenszeiten und Status-Codes 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

Abbildung 13-5
Dialogfeld "Kaplan-Meier: Optionen"



Sie können verschiedene Arten von Ausgabe für die Kaplan-Meier-Analyse anfordern.

Statistiken. 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.

- **Überlebenswahrscheinlichkeit.** Zeigt die kumulative Überlebensfunktion auf einer linearen Skala an.
- **Eins minus Überleben.** Erzeugt ein Diagramm der Werte "1 - Überlebensfunktion" auf einer linearen Skala.

- **Kumulative Hazardrate.** Zeigt die kumulative Hazard-Funktion auf einer linearen Skala an.
- **Log-Überleben.** Zeigt kumulative Überlebensfunktion auf einer logarithmischen Skala an.

Zusätzliche Funktionen beim Befehl KM

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Anfordern von Häufigkeitstabellen, bei denen nachzutragende fehlende Fälle als eine von zensierten Fällen separate Kategorie berücksichtigt werden.
- Festlegen von ungleichen Abständen für den Test auf lineare Trends.
- Berechnen von Perzentilen für die Überlebenszeitvariable, die nicht Quartilen entsprechen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Cox-Regressionsanalyse

Cox Regression builds a predictive model for time-to-event data. Das Modell erzeugt eine Überlebensfunktion, die die Wahrscheinlichkeit vorhersagt, mit der das interessierende Ereignis zu einer gegebenen Zeit t für vorgegebene Werte der Einflussvariablen aufgetreten ist. Die Form der Überlebensfunktion und die Regressionskoeffizienten für die Einflussvariablen werden aus beobachteten Fällen geschätzt. Das Modell kann dann auf neue Fälle angewendet werden, für die Messungen der Einflussvariablen 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.

Statistiken. Für jedes Modell: $-2LL$, die Likelihood-Quotienten-Statistik und das Chi-Quadrat insgesamt. Für die Variablen im Modell: Parameterschätzer, Standardfehler und Wald-Statistiken. Für Variablen, die nicht im Modell enthalten sind: Statistiken der Werte und Residuen-Chi-Quadrat.

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 indikatorkodierte Variablen sein (die Prozedur enthält eine Option zum automatischen Umkodieren für kategoriale Variablen). Schichtvariablen müssen kategorial und als ganze Zahlen oder als kurze Strings kodiert sein.

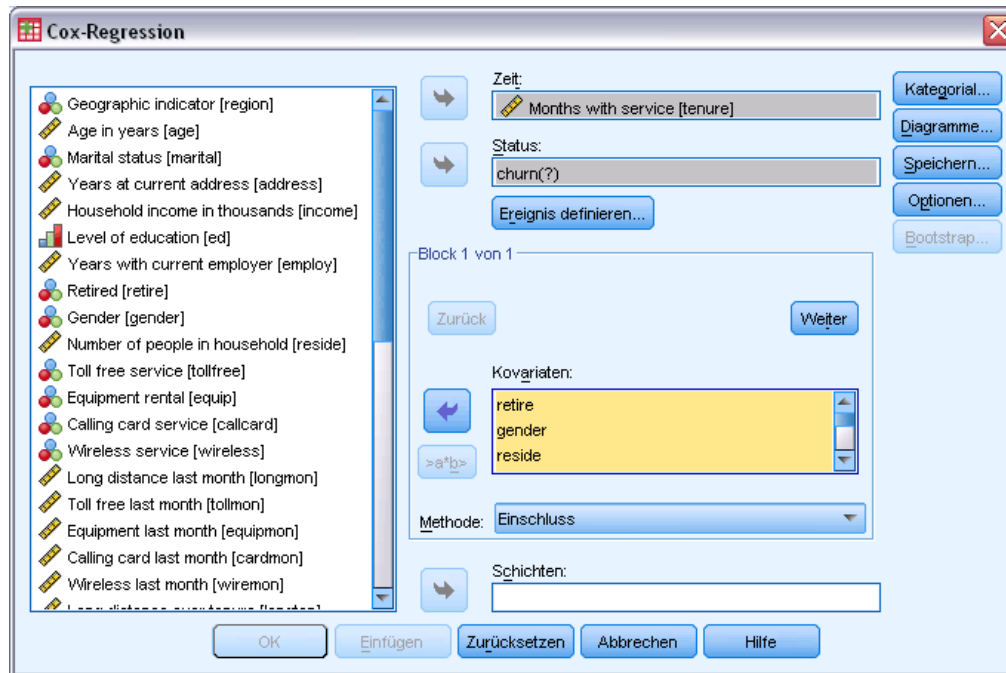
Annahmen. Beobachtungen müssen unabhängige Variablen sein und das Hazard-Verhältnis muss über die Zeit konstant sein. Das Hazard-Verhältnis von einem Fall zum nächsten darf sich über die Zeit also nicht ändern. Letztere Annahme ist als die **Annahme über den Hazard-Anteil** bekannt.

Verwandte Prozeduren. Falls die Annahme über den Hazard-Anteil 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 Hazard-Funktionen für die Stichprobe(n) zu untersuchen. Falls in der Stichprobe keine zensierten Daten vorhanden sind (d. h., in jedem Fall das endgültige Ereignis eintrat), können Sie die Prozedur "Lineare Regression" verwenden, um das Verhältnis zwischen Einflußvariablen und Zeitpunkt bis zum Eintreten des Ereignisses zu modellieren.

Berechnen einer Cox-Regressionsanalyse

- Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Überleben > Cox-Regression...

Abbildung 14-1
Dialogfeld "Cox-Regression"



- Wählen Sie eine Zeitvariable aus. Fälle, deren Zeitwerte negativ sind, werden nicht analysiert.
- Wählen Sie eine Statusvariable aus und klicken Sie anschließend auf Ereignis definieren.
- Wählen Sie mindestens eine Kovariante aus. Um alle Wechselwirkungsterme einzuschließen, wählen Sie erst alle mit der Wechselwirkung verbundenen Variablen aus und klicken Sie dann auf >a*b>.

Wahlweise können Sie getrennte Modelle für unterschiedliche Gruppen berechnen, indem Sie eine Schichtvariable definieren.

Cox-Regression: Kategoriale Kovariaten definieren

Abbildung 14-2
Dialogfeld "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 Wechselwirkung in einer beliebigen Schicht. Falls einige der Variablen String-Variablen oder kategoriale Variablen sind, können Sie diese nur als kategoriale Kovariaten verwenden.

Kategoriale Kovariaten.Enthält die Variablen, die als kategoriale festgelegt wurden. Bei jeder Variablen wird in Klammern die zu verwendende Kontrastkodierung angezeigt. String-Variablen (durch das Symbol < nach deren 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 Einflußvariablen 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 Einflußvariablen mit Ausnahme der letzten Kategorie wird mit dem durchschnittlichen Effekt der nachfolgenden Kategorien verglichen.
- **Wiederholt.**Jede Kategorie der Einflußvariablen 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 Einflussvariablen 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 String-Kovariaten müssen kategoriale Kovariaten sein. Um eine String-Variable aus der Liste "Kategoriale Kovariaten" zu entfernen, müssen Sie alle Terme, welche diese Variable enthalten, aus der Liste "Kovariaten" im Hauptdialogfeld entfernen.

Diagramme für Cox-Regressionen

Abbildung 14-3
Dialogfeld "Cox-Regression: Diagramme"

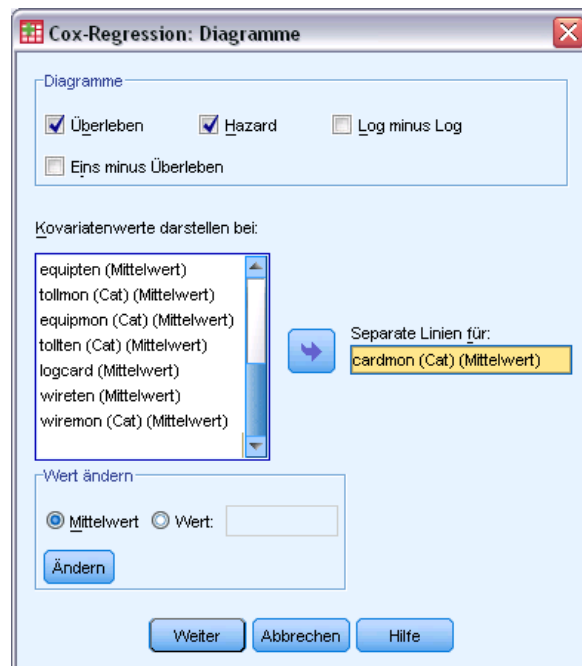


Diagramme können ein Hilfsmittel zur Bewertung des geschätzten Modells und zur Interpretation der Ergebnisse sein. Die Überlebens-, Hazard- und Log-Minus-Log-Funktionen sowie Eins minus Überleben können grafisch dargestellt werden.

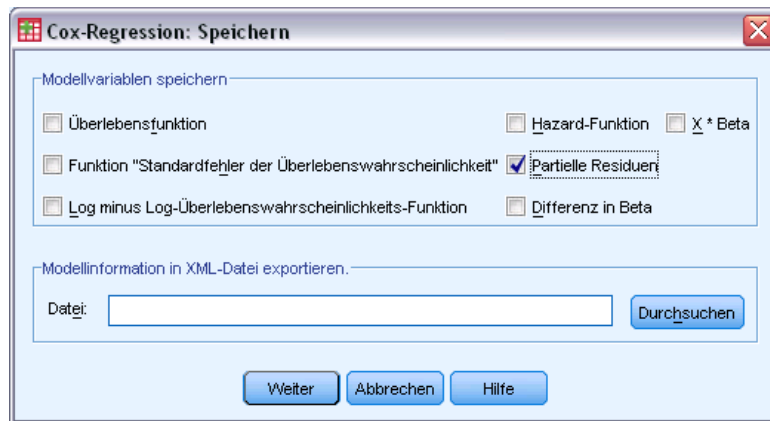
- **Überleben.** Zeigt die kumulative Überlebensfunktion auf einer linearen Skala an.
- **Hazard.** Zeigt die kumulative Hazard-Funktion auf einer linearen Skala an.
- **Log-minus-Log.** Die kumulative Überlebensschätzung nach Anwendung der $\ln(-\ln)$ -Transformation auf den Schätzer.
- **Eins minus Überleben.** Erzeugt ein Diagramm der Werte "1 - Überlebensfunktion" auf einer linearen Skala.

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.

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.

Cox-Regression: Neue Variablen speichern

Abbildung 14-4
Dialogfeld "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, Hazard-Funktion, partielle Residuen, DfBeta für die Regression und die lineare Einflussvariable $X \cdot \text{Beta}$ als neue Variablen zu speichern.

- **Überlebensfunktion.** Der Wert der kumulierten Überlebensfunktion für eine bestimmte Zeit. Er entspricht der Überlebenswahrscheinlichkeit bis zu dieser Zeit.
- **Log minus Log-Überlebenswahrscheinlichkeits-Funktion.** Die kumulierte Überlebensschätzung nach Anwendung der $\ln(-\ln)$ -Transformation auf den Schätzer.
- **Hazard-Funktion.** Speichert den Schätzer für die kumulative Hazard-Funktion (auch Cox-Snell-Residuum genannt).
- **Partielle Residuen.** Sie können partielle Residuen über Überlebenszeit auftragen, um die proportionale Hazard-Annahme zu prüfen. Für jede Kovariante im endgültigen Modell wird eine Variable gespeichert. Partielle Residuen stehen nur für Modelle mit mindestens einer Kovariante zur Verfügung.

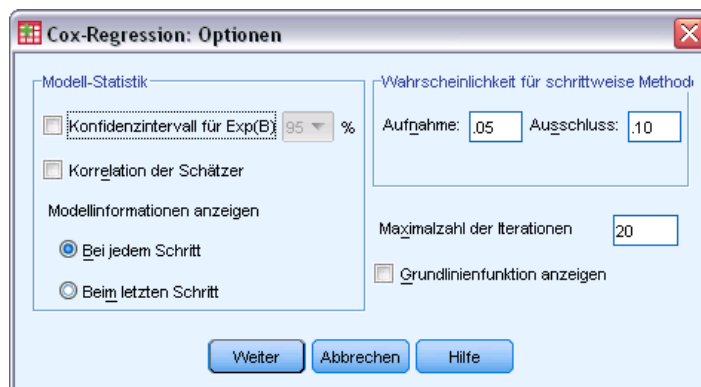
- **DfBeta.** Die geschätzte Änderung bei einem Koeffizienten, wenn ein Fall ausgeschlossen wird. Für jede Kovariante im endgültigen Modell wird eine Variable gespeichert. Differenzen in Beta stehen nur für Modelle mit mindestens einer Kovariante zur Verfügung.
- **X*Beta.** Linearer Einflusswert. Linearer Einflußwert. Die Summe aus dem Produkt mittelwertzentrierter Kovariatenwerte und deren jeweiligen Parameterschätzern für jeden Fall.

Falls Sie die Cox-Regression mit einer zeitabhängigen Kovariaten verwenden, sind DfBeta und die lineare Einflussvariable X*Beta die einzigen Variablen, die Sie speichern können.

Modellinformation in XML-Datei exportieren. Parameterschätzer werden im XML-Format in die angegebene Datei exportiert. Anhand dieser Modelldatei können Sie die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Cox-Regression: Optionen

Abbildung 14-5
Dialogfeld "Cox-Regression: Optionen"



Sie können verschiedene Aspekte der Analyse und Ausgabe steuern.

Modell-Statistik. Für die Modellparameter stehen Statistiken wie Konfidenzintervalle für $\text{Exp}(B)$ und Korrelation der Schätzer zur Verfügung. Diese Statistiken können für jeden Schritt oder nur für den letzten Schritt angefordert werden.

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 Signifikanzniveau größer als der Ausschlusswert ist. Der Aufnahmewert muss kleiner als der Ausschlusswert sein.

Maximalzahl der Iterationen. Hiermit können Sie die Maximalzahl der Iterationen für das Modell festlegen, die bestimmt, wie lange die Prozedur nach einer Lösung sucht.

Grundlinienfunktion anzeigen. Hiermit können Sie die Hazard-Grundlinienfunktion 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 Endereignis 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

Mit der Befehlssyntax können Sie auch Folgendes:

- Anfordern von Häufigkeitstabellen, bei denen nachzutragende fehlende Fälle als eine von zensierten Fällen separate Kategorie berücksichtigt werden.
- Auswählen einer Referenzkategorie für die Methoden der Abweichungskontraste, einfachen Kontraste und Indikatorkontraste, die sich von der ersten und letzten Kategorie unterscheidet.
- Festlegen von ungleichen Kategorienabständen für die polynomiale Kontrastmethode.
- Festlegen von zusätzlichen Iterationskriterien.
- Sie können festlegen, wie mit fehlenden Werten verfahren wird.
- Festlegen der Namen für gespeicherte Variablen.
- Schreiben der Ausgabe in eine externe IBM® SPSS® Statistics-Systemdatei.
- Zwischenspeichern der Daten für jede Gruppe einer aufgeteilten Datei während der Verarbeitung in einer externen Hilfsdatei. Dadurch lassen sich beim Ausführen von Analysen mit großen Daten-Sets Speicherkapazitäten freihalten. Bei zeitabhängigen Kovariaten steht diese Möglichkeit nicht zur Verfügung.

Siehe *Befehlssyntaxreferenz* für die vollständigen Syntaxinformationen.

Berechnen von zeitabhängigen Kovariaten

Beim Berechnen von Cox-Regressionsmodellen gibt es Situationen, in denen die proportionalen Hazard-Annahmen nicht erfüllt werden. Dies können Änderungen der Hazard-Verhältnisse über Zeit sein, wenn also die Werte einer (oder mehrerer) Kovariaten zu verschiedenen Zeitpunkten unterschiedlich sind. In diesen Fällen ist es notwendig, dass Sie ein erweitertes Cox-Regressionsmodell verwenden, in dem Sie **zeitabhängige Kovariaten** festlegen können.

Um ein solches Modell zu analysieren, müssen Sie zuerst die zeitabhängige Kovariate definieren. (Mehrere zeitabhängige Kovariaten können mit der Befehlssyntax festgelegt werden.) Zur Erleichterung hierbei steht eine Systemvariable zur Verfügung, welche die Zeit darstellt. Diese Variable wird mit $T_$ bezeichnet. Sie können diese Variable verwenden, um zeitabhängige Kovariaten auf zwei Arten zu definieren.

- Wenn Sie die proportionale Hazard-Annahme in Bezug auf eine bestimmte Kovariate testen oder ein erweitertes Cox-Regressionsmodell schätzen möchten, mit dem nichtproportionale Hazards möglich sind, definieren Sie die zeitabhängige Kovariate als eine Funktion der Zeitvariablen $T_$ und der fraglichen Kovariate. Ein geläufiges Beispiel wäre das einfache Produkt aus Zeitvariable und Kovariate, aber es können auch komplexere Funktionen festgelegt werden. Durch Testen der Signifikanz des Koeffizienten der zeitabhängigen Kovariate können Sie feststellen, ob die proportionale Hazard-Annahme angemessen ist.
- Einige Variablen können zu unterschiedlichen Zeiten unterschiedliche Werte aufweisen, weisen aber kein systematisches Verhältnis zur Zeit auf. In solchen Fälle müssen Sie eine **segmentierte zeitabhängige Kovariate** definieren. Dies können sie mit **logischen Ausdrücken** erreichen. 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 Satz von Messwerten erstellen. Wenn Sie zum Beispiel über vier Wochen einer Studie einmal wöchentlich den Blutdruck gemessen haben (gekennzeichnet durch $BP1$ bis $BP4$), können Sie die zeitabhängige Kovariate durch den folgenden Ausdruck definieren: $(T_ < 1) * BP1 + (T_ >= 1 \ \& \ T_ < 2) * BP2 + (T_ >= 2 \ \& \ T_ < 3) * BP3 + (T_ >= 3 \ \& \ T_ < 4) * BP4$. Beachten Sie, dass bei einem gegebenen Fall genau einer der Ausdrücke in Klammern dem Wert 1 entspricht; alle anderen Ausdrücke in Klammern weisen den Wert 0 auf. Diese Funktion kann folgendermaßen interpretiert werden: Wenn die Zeitspanne kürzer als eine Woche ist, wird $BP1$ verwendet, wenn die Zeitspanne länger als eine Woche aber kürzer als zwei Wochen ist, wird $BP2$ verwendet und so weiter.

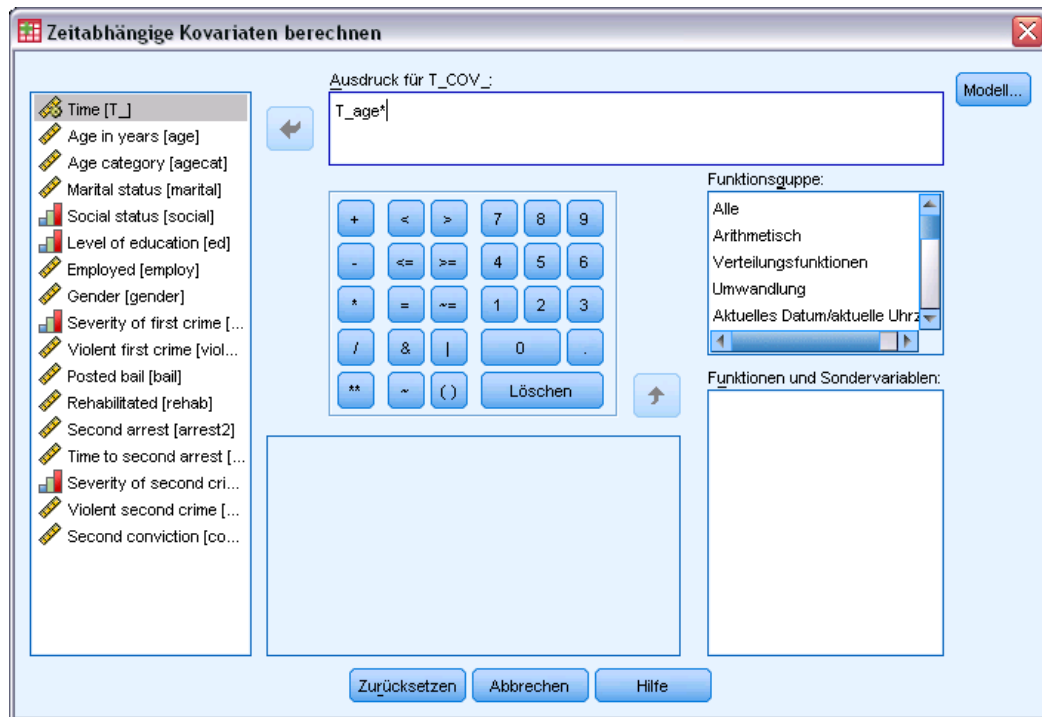
Sie können die Steuerelemente im Dialogfeld "Zeitabhängige Kovariaten berechnen" verwenden, um den Ausdruck für die zeitabhängige Kovariate zu bilden, oder Sie können diesen direkt in das Eingabefeld "Ausdruck für $T_COV_$ " eingeben. Beachten Sie, dass String-Konstanten in Anführungszeichen oder Apostrophe gesetzt und numerische Konstanten in amerikanischem Format mit einem Punkt als Dezimaltrennzeichen eingegeben werden

müssen. Die resultierende Variable wird als $T_COV_$ bezeichnet und sollte als Kovariate in das Cox-Regressionsmodell eingeschlossen werden.

Berechnen einer zeitabhängigen Kovariate

- Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Überlebenswahrscheinlichkeit > Cox mit zeitabh. Kovariaten...

Abbildung 15-1
Dialogfeld "Zeitabhängige Kovariaten berechnen"



- Geben Sie einen Ausdruck für die zeitabhängige Kovariate ein.
- Wählen Sie Modell aus, um mit der Cox-Regression fortzufahren.

Anmerkung: Vergewissern Sie sich, dass Sie die neue Variable $T_COV_$ als Kovariate in das Cox-Regressionsmodell aufgenommen haben.

Für weitere Informationen siehe Thema Cox-Regressionsanalyse in Kapitel 14 auf S. 154.

Zusätzliche Funktionen bei der Cox-Regression mit zeitabhängigen Kovariaten

Mit der Befehlssyntax-Sprache können Sie auch mehrere zeitabhängige Kovariaten festlegen. Außerdem stehen weitere Funktionen der Befehlssyntax für die Cox-Regression mit oder ohne zeitabhängige Kovariaten zur Verfügung.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kodierungsschemata 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 kodiert werden soll (in der Regel im Unterbefehl `CONTRAST`). Dieser Anhang erläutert die verschiedenen Kontrasttypen des Unterbefehls `CONTRAST`.

Abweichung

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

Mittelwert	($1/k$	$1/k$...	$1/k$	$1/k$)
df(1)	($1-1/k$	$-1/k$...	$-1/k$	$-1/k$)
df(2)	($-1/k$	$1-1/k$...	$-1/k$	$-1/k$)
.			.			
.			.			
df(k-1)	($-1/k$	$-1/k$...	$1-1/k$	$-1/k$)

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

($1/3$	$1/3$	$1/3$)
($2/3$	$-1/3$	$-1/3$)
($-1/3$	$2/3$	$-1/3$)

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

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

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

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & -1/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$$

Einfach

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

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

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:

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

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

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

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

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

Helmert

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

Mittelwert	($1/k$	$1/k$...	$1/k$	$1/k$)
df(1)	(1	$-1/(k-1)$...	$-1/(k-1)$	$-1/(k-1)$)
df(2)	(0	1	...	$-1/(k-2)$	$-1/(k-2)$)
.
df(k-2)	(0	0	1	$-1/2$	$-1/2$)
df(k-1)	(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:

Mittelwert	($1/k$	$1/k$	$1/k$...	$1/k$)
df(1)	(-1	1	0	...	0)
df(2)	($-1/2$	$-1/2$	1	...	0)
.
df(k-1)	($-1/(k-1)$	$-1/(k-1)$	$-1/(k-1)$...	1)

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 (WIRKSTF) =POLYNOMIAL
```

dem Unterbefehl

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

Gleiche Abstände sind jedoch nicht immer erforderlich. Angenommen, *Wirkstf* stellt verschiedene Dosierungen eines Wirkstoffs dar, der drei verschiedenen Gruppen verabreicht wurde. Wenn der zweiten Gruppe eine doppelt so hohe Dosierung wie der ersten Gruppe und der dritten Gruppe eine dreimal so hohe Dosierung wie der ersten Gruppe verabreicht wurde, weisen die Behandlungskategorien gleiche Abstände auf. Hierbei eignet sich die folgende, aus aufeinander folgenden Ganzzahlen bestehende Metrik:

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

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

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

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

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

Wiederholt

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

Mittelwert	($1/k$	$1/k$	$1/k$...	$1/k$	$1/k$)
df(1)	(1	-1	0	...	0	0)
df(2)	(0	1	-1	...	0	0)

$$df(k-1) \quad (0 \quad 0 \quad 0 \quad \dots \quad 1 \quad -1)$$

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:

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

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

Speziell (GLM Ptable)

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 Mittelwert-Effekt (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 gewünschten 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:

$(1 \quad 1 \quad 1 \quad 1)$	Gewichtungen für Berechnung des Mittelwerts
$(3 \quad -1 \quad -1 \quad -1)$	Vergleich 1. mit 2. bis 4.
$(0 \quad 2 \quad -1 \quad -1)$	Vergleich 2. mit 3. und 4.
$(0 \quad 0 \quad 1 \quad -1)$	Vergleich 3. mit 4.

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 )
```

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

$$\text{Zeilen 2 und 3:} \quad (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0$$

$$\text{Zeilen 2 und 4:} \quad (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0$$

$$\text{Zeilen 3 und 4:} \quad (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0$$

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

Kodierung für Indikatorvariablen. Auch bekannt als Dummy-Kodierung. Dieser Kontrast ist bei LOGLINEAR und MANOVA nicht verfügbar. Die Anzahl der neu kodierten Variablen entspricht $k-1$. Fälle in der Referenzkategorie werden für alle $k-1$ Variablen als 0 kodiert. Ein Fall in der i -ten Kategorie wird für fast alle Indikatorvariablen als 0 und lediglich für die i -ten als 1 kodiert.

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 ergibt sich als Produkt der Korrelationen zwischen den Elementen, die dazwischen liegen.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3 \\ \sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}$$

AR(1). Diese Struktur ist autoregressiv erster Ordnung mit homogenen Varianzen. Die Korrelation zwischen zwei beliebigen Elementen ist gleich ρ für benachbarte Elemente, ρ^2 für Elemente, die durch ein drittes Element getrennt sind, usw. ρ ist beschränkt, sodass $-1 < \rho < 1$.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

AR(1): heterogen. Diese Struktur ist autoregressiv erster Ordnung mit heterogenen Varianzen. Die Korrelation zwischen zwei beliebigen Elementen ist gleich ρ für benachbarte Elemente, ρ^2 für zwei Elemente, die durch ein drittes Element getrennt sind, usw. ρ ist beschränkt und muss zwischen -1 und 1 liegen.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

ARMA(1,1). Diese Struktur ist autoregressiv mit gleitendem Durchschnitt erster Ordnung. Sie weist homogene Varianz auf. Die Korrelation zwischen zwei Elementen ist $\phi^*\rho$ für benachbarte Elemente, $\phi^*(\rho^2)$ für Elemente, die durch ein drittes Element getrennt sind, usw. Dabei sind ρ und ϕ die Parameter des autoregressiven Anteils bzw. des gleitenden Durchschnittsanteils; ihre Werte müssen zwischen -1 und 1 liegen.

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi\rho^2 \\ \phi\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & 1 \end{bmatrix}$$

Zusammengesetzt symmetrisch (ZS). Diese Struktur hat eine konstante Varianz und eine konstante Kovarianz.

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

Zusammengesetzt symmetrisch: Korrelationsmetrik. Kovarianzstruktur mit homogenen Varianzen und homogenen Korrelationen zwischen den Elementen.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Zusammengesetzt symmetrisch: heterogen. Kovarianzstruktur mit heterogenen Varianzen und konstanten Korrelationen zwischen den Elementen.

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

Diagonal. Kovarianzstruktur mit heterogenen Varianzen und Korrelation zwischen den Elementen gleich Null.

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

Faktor-analytisch: 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 Varianzteile.

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

Faktor-analytisch: 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 Wurzel aus dem Produkt der ersten ihrer beiden heterogenen Varianzanteile.

$$\begin{bmatrix} \lambda_1^2 + d_1 & \lambda_2 \lambda_1 & \lambda_3 \lambda_1 & \lambda_4 \lambda_1 \\ \lambda_2 \lambda_1 & \lambda_2^2 + d_2 & \lambda_3 \lambda_2 & \lambda_4 \lambda_2 \\ \lambda_3 \lambda_1 & \lambda_3 \lambda_2 & \lambda_3^2 + d_3 & \lambda_4 \lambda_3 \\ \lambda_4 \lambda_1 & \lambda_4 \lambda_2 & \lambda_4 \lambda_3 & \lambda_4^2 + d_4 \end{bmatrix}$$

Huynh-Feldt. Dies ist eine "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.

$$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda & \sigma_4^2 \end{bmatrix}$$

Skalierte Identität. Diese Struktur hat eine konstante Varianz. Es wird angenommen, dass es zwischen den Elementen keine Korrelationen gibt.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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.

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

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.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_1 & \sigma_3 \sigma_1 \rho_2 & \sigma_4 \sigma_1 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_3 \sigma_2 \rho_1 & \sigma_4 \sigma_2 \rho_2 \\ \sigma_3 \sigma_1 \rho_2 & \sigma_3 \sigma_2 \rho_1 & \sigma_3^2 & \sigma_4 \sigma_3 \rho_1 \\ \sigma_4 \sigma_1 \rho_3 & \sigma_4 \sigma_2 \rho_2 & \sigma_4 \sigma_3 \rho_1 & \sigma_4^2 \end{bmatrix}$$

Unstrukturiert (UN). Eine ganz allgemeine Kovarianzmatrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Unstrukturiert: Korrelationsmetrik. Kovarianzstruktur mit heterogenen Varianzen und heterogenen Korrelationen.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$$

Varianzkomponenten. Damit wird jedem der angegebenen Zufallseffekte eine skalierte Identitätsstruktur (ID) zugewiesen.

Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing,

using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



- A-priori-Zufallseffekte
 - in Varianzkomponenten, 35
- Abweichungsresiduen
 - in: Verallgemeinerte lineare Modelle, 72
- Allgemeine loglineare Analyse
 - Anzeigeoptionen, 134
 - Faktoren, 131
 - Grafik, 134
 - Konfidenzintervalle, 134
 - Kontraste, 131
 - Kriterien, 134
 - Modelleinstellungen, 133
 - Residuen, 135
 - Speichern von Variablen, 135
 - Speichern vorhergesagter Werte, 135
 - Verteilung der Zellenhäufigkeiten, 131
 - Zellenkovariaten, 131
 - Zellenstrukturen, 131
 - zusätzliche Funktionen beim Befehl, 135
- Allgemeine schätzbare Funktion
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Analyse der Kovarianz
 - in GLM - Multivariat, 2
- ANOVA (Variance Components Options)
 - in GLM - Messwiederholungen, 16
 - in GLM - Multivariat, 2
- Ausgeschlossene Residuen
 - in GLM, 11
 - in GLM - Messwiederholungen, 28

- Bartlett-Test auf Sphärizität
 - in GLM - Multivariat, 13
- Benutzerdefinierte Modelle
 - in GLM - Messwiederholungen, 21
 - in “Modell für loglineare Analyse auswählen”, 128
 - in Varianzkomponenten, 34
- Beobachtete Mittelwerte
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Bewertung
 - in “Lineare gemischte Modelle”, 46
- Binomialverteilung
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Binominalverteilung
 - in: Verallgemeinerte lineare Modelle, 54
- Bonferroni
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Box’ M-Test
 - in GLM - Multivariat, 13

- Breslow-Test
 - in Kaplan-Meier, 150

- C nach Dunnett
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Cook-Distanz
 - in GLM, 11
 - in GLM - Messwiederholungen, 28
 - in: Verallgemeinerte lineare Modelle, 71
- Cox-Regression, 154
 - Beispiel, 154
 - DfBeta, 158
 - Ereignis definieren, 160
 - Grafik, 157
 - Grundlinienfunktionen, 159
 - Hazard-Funktion, 158
 - Iteration, 159
 - Kategoriale Kovariaten, 156
 - Kontraste, 156
 - Kovariaten, 154
 - partielle Residuen, 158
 - schrittweise Aufnahme und Ausschluss, 159
 - Speichern von neuen Variablen, 158
 - Statistik, 154, 159
 - Statusvariable für Überlebensverteilung, 160
 - String-Kovariaten, 156
 - Überlebensfunktion, 158
 - zeitabhängige Kovariaten, 161–162
 - zusätzliche Funktionen beim Befehl, 160

- Deskriptive Statistik
 - in GLM - Messwiederholungen, 29
- Deskriptive Statistiken
 - in GLM - Multivariat, 13
 - in “Lineare gemischte Modelle”, 47
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Diagramme mit der Streubreite gegen das mittlere Niveau
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Duncans multipler Spannweitentest
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Dunnett-T-Test
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9

- Ehrlich signifikante Differenz nach Tukey
 - in GLM - Messwiederholungen, 26

- in GLM - Multivariat, 9
- Eingeschränkte Maximum-Likelihood-Schätzung
 - in Varianzkomponenten, 35
- Eta-Quadrat
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- F* nach R-E-G-W
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Faktoren
 - in GLM - Messwiederholungen, 20
- Faktorstufeninformationen
 - in "Lineare gemischte Modelle", 47
- feste Effekte
 - in "Lineare gemischte Modelle", 42
- Feste vorhergesagte Werte
 - in "Lineare gemischte Modelle", 50
- Fisher-Bewertung
 - in "Lineare gemischte Modelle", 46
- Gammaverteilung
 - in: Verallgemeinerte lineare Modelle, 54
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Gehan-Test
 - in Sterbetafeln, 146
- Gemischte Modelle
 - linear, 38
- Generalized Estimating Equations, 75
 - Anfangswerte, 91
 - Antwort, 83
 - Einflussvariablen, 85
 - Geschätzte Randmittel, 95
 - Modelleinstellungen, 87
 - Modellexport, 99
 - Optionen für kategoriale Faktoren, 86
 - Referenzkategorie für binäre Antwort, 84
 - Schätzkriterien, 89
 - Speichern von Variablen in der Arbeitsdatei, 97
 - Statistik, 93
 - Typ des Modells, 79
- Generalized Linear Models, 52
 - Antwort, 57
 - Einflussvariablen, 59
 - Geschätzte Randmittel, 68
 - Modelleinstellungen, 61
 - Modellexport, 72
 - Modelltypen, 52
 - Schätzkriterien, 63
 - Speichern von Variablen in der Arbeitsdatei, 70
 - Statistik, 66
 - Verknüpfungsfunktion, 52
 - Verteilung, 52
- Geringste signifikante Differenz
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Gesättigte Modelle
 - in GLM - Messwiederholungen, 21
 - in "Modell für loglineare Analyse auswählen", 128
 - in Varianzkomponenten, 34
- geschätzte Randmittel
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
 - in "Lineare gemischte Modelle", 49
- Geschätzte Randmittel
 - in: Verallgemeinerte lineare Modelle, 68
 - in: Verallgemeinerte Schätzungsgleichungen, 95
- Gewichtete Schätzwerte
 - in GLM, 11
 - in GLM - Messwiederholungen, 28
- GLM
 - Speichern von Matrizen, 11
 - Speichern von Variablen, 11
- GLM - Messwiederholungen, 16
 - anzeigen, 29
 - Diagnose, 29
 - Faktoren definieren, 20
 - geschätzte Randmittel, 29
 - Modell, 21
 - Optionen, 29
 - Post-Hoc-Tests, 26
 - Profilplots, 25
 - Speichern von Variablen, 28
 - zusätzliche Funktionen beim Befehl, 31
- GLM - Multivariat, 2, 14
 - Abhängige Variable, 2
 - anzeigen, 13
 - Diagnose, 13
 - Faktoren, 2
 - geschätzte Randmittel, 13
 - Kovariaten, 2
 - Optionen, 13
 - Post-Hoc-Tests, 9
 - Profilplots, 8
- GLOR
 - in "Allgemeine loglineare Analyse", 131
- Grafik
 - in "Allgemeine loglineare Analyse", 134
 - in "Logit-loglineare Analyse", 140
- GT2 nach Hochberg
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Güte der Anpassung;Anpassungsgüte
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Häufigkeiten
 - in "Modell für loglineare Analyse auswählen", 129
- Hazard-Rate
 - in Sterbetafeln, 143
- Hebelwerte
 - in GLM, 11
 - in GLM - Messwiederholungen, 28

- in: Verallgemeinerte lineare Modelle, 71
- Hierarchische loglineare Modelle, 126
- Hierarchische Zerlegung, 5, 22
 - in Varianzkomponenten, 36
- Identitätsverknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- inverse Normalverteilung
 - in: Verallgemeinerte lineare Modelle, 54
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Iteration
 - in “Modell für loglineare Analyse auswählen”, 129
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89
- Iterationsprotokoll
 - in “Lineare gemischte Modelle”, 46
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Kaplan-Meier, 148
 - Beispiel, 148
 - Definieren von Ereignissen, 150
 - Grafik, 152
 - linearer Trend für Faktorstufen, 150
 - Mittelwert und Median der Überlebenszeit, 152
 - Quartile, 152
 - Speichern von neuen Variablen, 151
 - Statistik, 148, 152
 - Statusvariablen für Überlebensverteilung, 150
 - Überlebensstabellen, 152
 - Vergleichen von Faktorstufen, 150
 - zusätzliche Funktionen beim Befehl, 153
- Konfidenzintervalle
 - in “Allgemeine loglineare Analyse”, 134
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
 - in “Lineare gemischte Modelle”, 47
 - in “Logit-loglineare Analyse”, 140
- Konstruieren von Termen, 5, 22, 34, 128, 133, 140
- Kontingenztafeln
 - in “Allgemeine loglineare Analyse”, 131
- Kontraste
 - in “Allgemeine loglineare Analyse”, 131
 - in Cox-Regression, 156
 - in “Logit-loglineare Analyse”, 137
- Kontrastkoeffizienten-Matrix
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Konvergenz der Hesse-Matrix
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89
- Korrelationsmatrix
 - in “Lineare gemischte Modelle”, 47
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Kovarianz-Parametertest
 - in “Lineare gemischte Modelle”, 47
- Kovarianzmatrix
 - in GLM, 11
 - in “Lineare gemischte Modelle”, 47
 - in: Verallgemeinerte lineare Modelle, 63, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 89, 94
- Kovarianzmatrix der Zufallseffekte
 - in “Lineare gemischte Modelle”, 47
- Kovarianzstrukturen, 169
 - in “Lineare gemischte Modelle”, 169
- Kovariaten
 - in Cox-Regression, 156
- Kreuzprodukte
 - Hypothesen- und Fehlermatrizen, 13
- Kreuztabelle
 - in “Modell für loglineare Analyse auswählen”, 126
- kumulative Cauchit-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- kumulative Logit-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- kumulative Probit-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- L-Matrix
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Lagrange-Multiplikator-Test
 - in: Verallgemeinerte lineare Modelle, 67
- legal notices, 173
- Levene-Test
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Likelihood-Residuen
 - in: Verallgemeinerte lineare Modelle, 72
- Lineare gemischte Modelle, 38, 169
 - feste Effekte, 42
 - geschätzte Randmittel, 49
 - Konstruieren von Termen, 42–43
 - Kovarianzstruktur, 169
 - Modell, 47
 - Schätzkriterien, 46
 - Variablen speichern, 50
 - Wechselwirkungsterme, 42
 - Zufällige Effekte, 44
 - zusätzliche Funktionen beim Befehl, 50
- Log-Likelihood-Konvergenz
 - in “Lineare gemischte Modelle”, 46
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89
- Log-Rang-Test
 - in Kaplan-Meier, 150
- Log-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55

- in: Verallgemeinerte Schätzungsgleichungen, 81
- Logit-loglineare Analyse, 137
 - Anzeigoptionen, 140
 - Faktoren, 137
 - Grafik, 140
 - Konfidenzintervalle, 140
 - Kontraste, 137
 - Kriterien, 140
 - Modelleinstellungen, 139
 - Residuen, 141
 - Speichern von Variablen, 141
 - Verteilung der Zellenhäufigkeiten, 137
 - Vorhergesagte Werte, 141
 - Zellenkovariaten, 137
 - Zellenstrukturen, 137
- Logit-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- Loglineare Analyse, 126
 - Allgemeine loglineare Analyse, 131
 - Logit-loglineare Analyse, 137
- LSD nach Fisher
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Mauchly-Test auf Sphärizität
 - in GLM - Messwiederholungen, 29
- Maximum-Likelihood-Schätzung
 - in Varianzkomponenten, 35
- MINQUE (Variance Components Options)
 - in Varianzkomponenten, 35
- Modell für loglineare Analyse auswählen, 126
 - Definieren der Faktorenbereiche, 127
 - Modelle, 128
 - Optionen, 129
 - zusätzliche Funktionen beim Befehl, 130
- Modellansicht
 - in "Verallgemeinerte lineare gemischte Modelle", 119
- Modellbildende Klasse
 - in "Modell für loglineare Analyse auswählen", 128
- Modellinformationen
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Multinomiale Logit-Modelle, 137
- Multinomialverteilung
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Multinomialverteilung
 - in: Verallgemeinerte lineare Modelle, 54
- Multipler Spannweitentest nach Ryan-Einot-Gabriel-Welsch
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Multipler F nach Ryan-Einot-Gabriel-Welsch
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Multivariat: ANOVA, 2
- Multivariat: GLM, 2
- Multivariate Regression, 2
- negative Binomialverteilung
 - in: Verallgemeinerte lineare Modelle, 54
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Newman-Keuls
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Newton-Raphson-Methode
 - in "Allgemeine loglineare Analyse", 131
 - in "Logit-loglineare Analyse", 137
- Nicht standardisierte Residuen
 - in GLM, 11
 - in GLM - Messwiederholungen, 28
- Normalverteilung
 - in: Verallgemeinerte lineare Modelle, 54
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Normalverteilungsdiagramme
 - in "Modell für loglineare Analyse auswählen", 129
- Paarweiser Vergleichstest nach Gabriel
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Paarweiser Vergleichstest nach Games und Howell
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Parameter-Kovarianzmatrix
 - in "Lineare gemischte Modelle", 47
- Parameterkonvergenz
 - in "Lineare gemischte Modelle", 46
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89
- Parameterschätzer
 - in "Allgemeine loglineare Analyse", 131
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
 - in "Lineare gemischte Modelle", 47
 - in "Logit-loglineare Analyse", 137
 - in "Modell für loglineare Analyse auswählen", 129
 - in: Verallgemeinerte lineare Modelle, 67
 - in: Verallgemeinerte Schätzungsgleichungen, 94
- Pearson-Residuen
 - in: Verallgemeinerte lineare Modelle, 71
 - in: Verallgemeinerte Schätzungsgleichungen, 98
- Poisson-Regression
 - in "Allgemeine loglineare Analyse", 131
- Poissonverteilung
 - in: Verallgemeinerte lineare Modelle, 54
 - in: Verallgemeinerte Schätzungsgleichungen, 80
- Power-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- Probit-Verknüpfungsfunktion
 - in: Verallgemeinerte lineare Modelle, 55
 - in: Verallgemeinerte Schätzungsgleichungen, 81
- Profilplots
 - in GLM - Messwiederholungen, 25

- in GLM - Multivariat, 8
- Proportionales Hazard-Modell
 - in Cox-Regression, 154
- Q* nach R-E-G-W
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Quadratsumme, 5, 22
- Quadratsummen
 - Hypothesen- und Fehlermatrizen, 13
 - in "Lineare gemischte Modelle", 43
 - in Varianzkomponenten, 36
- Quadratsummen und Kreuzprodukte
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Quadratsummen und Kreuzprodukte für Residuen; Residuen: Quadratsummen und Kreuzprodukte
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Quotenverhältnis
 - in "Allgemeine loglineare Analyse", 131
- Referenzkategorie
 - in: Verallgemeinerte lineare Modelle, 58
 - in: Verallgemeinerte Schätzungsgleichungen, 84, 86
- Residuen
 - in "Allgemeine loglineare Analyse", 135
 - in "Lineare gemischte Modelle", 50
 - in "Logit-loglineare Analyse", 141
 - in "Modell für loglineare Analyse auswählen", 129
 - in: Verallgemeinerte lineare Modelle, 71
 - in: Verallgemeinerte Schätzungsgleichungen, 98
- Residuen-Diagramme
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Residuen-Kovarianzmatrix
 - in "Lineare gemischte Modelle", 47
- Rückwärtselimination
 - in "Modell für loglineare Analyse auswählen", 126
- Schätzer der Effektgröße
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Schätzer der Schärfe
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Scheffé-Test
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Schritt-Halbierungen
 - in "Lineare gemischte Modelle", 46
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89
- Segmentierte zeitabhängige Kovariaten
 - in Cox-Regression, 161
- Sidak-T-Test
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Skalenparameter
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89
- Standardabweichung
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Standardfehler
 - in GLM, 11
 - in GLM - Messwiederholungen, 28–29
 - in GLM - Multivariat, 13
- Standardisierte Residuen
 - in GLM, 11
 - in GLM - Messwiederholungen, 28
- Sterbetafel (Life Tables Options), 143
 - Anzeige von Tabellen unterdrücken, 146
 - Beispiel, 143
 - Faktorvariablen, 146
 - Grafik, 146
 - Hazard-Rate, 143
 - Statistik, 143
 - Statusvariablen für Überlebensverteilung, 145
 - Überlebensfunktion, 143
 - Vergleichen von Faktorstufen, 146
 - Wilcoxon-Test (Gehan-Test), 146
 - zusätzliche Funktionen beim Befehl, 147
- String-Kovariaten
 - in Cox-Regression, 156
- Student-Newman-Keuls-Prozedur
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Subjektvariablen
 - in "Lineare gemischte Modelle", 40
- t-Test
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- T2 nach Tamhane
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- T3 nach Dunnett
 - in GLM - Messwiederholungen, 26
 - in GLM - Multivariat, 9
- Tarone-Ware-Test
 - in Kaplan-Meier, 150
- Tests auf Homogenität der Varianzen
 - in GLM - Messwiederholungen, 29
 - in GLM - Multivariat, 13
- Toleranz für Prüfung auf Singularität
 - in "Lineare gemischte Modelle", 46
- trademarks, 174
- Trennung
 - in: Verallgemeinerte lineare Modelle, 63
 - in: Verallgemeinerte Schätzungsgleichungen, 89

- Tukey-*B*-Test
 in GLM - Messwiederholungen, 26
 in GLM - Multivariat, 9
- Tweedie-Verteilung
 in: Verallgemeinerte lineare Modelle, 54
 in: Verallgemeinerte Schätzungsgleichungen, 80
- Überlebensanalyse
 in Cox-Regression, 154
 in Kaplan-Meier, 148
 in Sterbetafeln, 143
 zeitabhängige Cox-Regression, 161
- Überlebensfunktion
 in Sterbetafeln, 143
- Variablen für Messwiederholungen
 in "Lineare gemischte Modelle", 40
- Varianzanalyse
 in Varianzkomponenten, 35
- Varianzkomponenten, 32
 Modell, 34
 Optionen, 35
 Speichern von Ergebnissen, 37
 zusätzliche Funktionen beim Befehl, 37
- Verallgemeinerte lineare gemischte Modelle, 102
 Analysegewichtung, 114
 Anmerkungen, 121–123
 Benutzerdefinierte Terme, 110
 Block für zufällige Effekte, 112
 Datenstruktur, 120
 feste Effekte, 108
 geschätzte Mittel, 123–124
 Geschätzte Randmittel, 116
 Klassifikationstabelle, 120
 Modellansicht, 119
 Modellexport, 118
 Modellzusammenfassung, 119
 Offset, 114
 Speichern von Feldern, 118
 Verknüpfungsfunktion, 105
 Vorhersage nach Beobachtung, 120
 Zielverteilung, 105
 Zufällige Effekte, 111
- Verallgemeinerte lineare Modelle
 Anfangswerte, 65
 Optionen für kategoriale Faktoren, 60
 Referenzkategorie für binäre Antwort, 58
- Verallgemeinertes Log-Quotenverhältnis
 in "Allgemeine loglineare Analyse", 131
- Verknüpfungsfunktion
 Verallgemeinerte lineare gemischte Modelle, 105
- Verknüpfungsfunktion "Log. Komplement"
 in: Verallgemeinerte lineare Modelle, 55
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verknüpfungsfunktion "negativ binomial"
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verknüpfungsfunktion "negativ binomial"
 in: Verallgemeinerte lineare Modelle, 55
- Verknüpfungsfunktion "Odds Power"
 in: Verallgemeinerte lineare Modelle, 55
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verknüpfungsfunktion vom Typ "Log-Log komplementär"
 in: Verallgemeinerte lineare Modelle, 55
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verknüpfungsfunktion vom Typ "Log-Log kumulativ komplementär"
 in: Verallgemeinerte lineare Modelle, 55
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verknüpfungsfunktion vom Typ "Log-Log kumulativ negativ"
 in: Verallgemeinerte lineare Modelle, 55
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verknüpfungsfunktion vom Typ "Log-Log negativ"
 in: Verallgemeinerte lineare Modelle, 55
 in: Verallgemeinerte Schätzungsgleichungen, 81
- Verschachtelte Terme
 in "Lineare gemischte Modelle", 43
 in: Verallgemeinerte lineare Modelle, 61
 in: Verallgemeinerte Schätzungsgleichungen, 87
- Vorhergesagte Werte
 in "Allgemeine loglineare Analyse", 135
 in "Lineare gemischte Modelle", 50
 in "Logit-loglineare Analyse", 141
- Wald-Statistik
 in "Allgemeine loglineare Analyse", 131
 in "Logit-loglineare Analyse", 137
- Waller-Duncan-*T*-Test
 in GLM - Messwiederholungen, 26
 in GLM - Multivariat, 9
- Wechselwirkungsterme, 5, 22, 34, 128, 133, 140
 in "Lineare gemischte Modelle", 42
- Wilcoxon-Test
 in Sterbetafeln, 146
- zensierte Fälle
 in Cox-Regression, 154
 in Kaplan-Meier, 148
 in Sterbetafeln, 143
- Zufällige Effekte
 in "Lineare gemischte Modelle", 44
- Zusammenfassung der Fallverarbeitung
 in: Verallgemeinerte lineare Modelle, 67
 in: Verallgemeinerte Schätzungsgleichungen, 94