

IBM SPSS Complex Samples 22

IBM

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 59 gelesen werden.

Produktinformation

Diese Ausgabe bezieht sich auf Version 22, Release 0, Modifikation 0 von IBM SPSS Statistics und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuausgabe geändert wird.

Inhaltsverzeichnis

Kapitel 1. Einführung in die Prozeduren von Complex Samples 1

Eigenschaften komplexer Stichproben	1
Verwendung der Prozeduren für komplexe Stichproben	2
Plandateien	2
Weiterführende Literatur	3

Kapitel 2. Stichprobenziehung mithilfe eines komplexen Designs 5

Erstellen eines neuen Stichprobenplans	5
Stichprobenassistent: Designvariablen	5
Baumstruktursteuerungen zur Navigation im Stichprobenassistenten	6
Stichprobenassistent: Methode der Stichprobenziehung	6
Stichprobenassistent: Stichprobenumfang	7
Ungleiche Umfänge definieren	8
Stichprobenassistent: Ausgabevariablen	8
Stichprobenassistent: Planübersicht	9
Stichprobenassistent: Stichprobe ziehen: Auswahloptionen	9
Stichprobenassistent: Stichprobe ziehen: Ausgabedateien	9
Stichprobenassistent: Fertigstellen	10
Bearbeiten eines bestehenden Stichprobenplans	10
Stichprobenassistent: Planübersicht	10
Ausführen eines bestehenden Stichprobenplans	11
Zusätzliche Funktionen bei den Befehlen CSPLAN und CSSELECT	11

Kapitel 3. Vorbereiten einer komplexen Stichprobe für die Analyse 13

Erstellen eines neuen Analyseplans	13
Analysevorbereitungsassistent: Designvariablen	13
Baumstruktursteuerungen zur Navigation im Analyseassistenten	14
Analysevorbereitungsassistent: Schätzmethode	14
Analysevorbereitungsassistent: Umfang	14
Ungleiche Umfänge definieren	15
Analysevorbereitungsassistent: Planübersicht	16
Analysevorbereitungsassistent: Fertigstellen	16
Bearbeiten eines bestehenden Analyseplans	16
Analysevorbereitungsassistent: Planübersicht	16

Kapitel 4. Plan für komplexe Stichproben 17

Kapitel 5. Häufigkeiten für komplexe Stichproben 19

Häufigkeiten für komplexe Stichproben: Statistiken	19
Komplexe Stichproben: Fehlende Werte	20
Komplexe Stichproben: Optionen	21

Kapitel 6. Deskriptive Statistiken für komplexe Stichproben 23

Deskriptive Statistiken für komplexe Stichproben: Statistiken	23
Deskriptive Statistiken für komplexe Stichproben: Fehlende Werte	24
Komplexe Stichproben: Optionen	24

Kapitel 7. Kreuztabellen für komplexe Stichproben 25

Kreuztabellen für komplexe Stichproben - Statistik	25
Komplexe Stichproben: Fehlende Werte	26
Komplexe Stichproben: Optionen	27

Kapitel 8. Verhältnisse für komplexe Stichproben 29

Verhältnisse für komplexe Stichproben: Statistiken	29
Verhältnisse für komplexe Stichproben: Fehlende Werte	30
Komplexe Stichproben: Optionen	30

Kapitel 9. Allgemeines lineares Modell für komplexe Stichproben 31

Allgemeines lineares Modell für komplexe Stichproben	32
Allgemeines lineares Modell für komplexe Stichproben: Statistik	32
Hypothesentests für komplexe Stichproben	33
Allgemeines lineares Modell für komplexe Stichproben: Geschätzte Mittelwerte	34
Allgemeines lineares Modell für komplexe Stichproben: Speichern	34
Allgemeines lineares Modell für komplexe Stichproben: Optionen	35
Zusätzliche Funktionen beim Befehl CSGLM	35

Kapitel 10. Logistische Regression für komplexe Stichproben 37

Logistische Regression für komplexe Stichproben: Referenzkategorie	37
Logistische Regression für komplexe Stichproben: Modell	38
Logistische Regression für komplexe Stichproben: Statistik	38
Hypothesentests für komplexe Stichproben	39
Logistische Regression für komplexe Stichproben: Odds-Verhältnis	40
Logistische Regression für komplexe Stichproben: Speichern	40
Logistische Regression für komplexe Stichproben: Optionen	41
Zusätzliche Funktionen beim Befehl CSLOGISTIC	41

Kapitel 11. Ordinale Regression für komplexe Stichproben 43

Ordinale Regression für komplexe Stichproben: Antwortwahrscheinlichkeiten. 44
Ordinale Regression für komplexe Stichproben: Modell 44
Ordinale Regression für komplexe Stichproben: Statistik. 45
Hypothesentests für komplexe Stichproben. 46
Ordinale Regression für komplexe Stichproben: Odds-Verhältnisse 46
Ordinale Regression für komplexe Stichproben: Speichern 47
Ordinale Regression für komplexe Stichproben: Optionen 47
Zusätzliche Funktionen beim Befehl CSORDINAL 48

Kapitel 12. Cox-Regression für komplexe Stichproben 49

Ereignis definieren 50
Prädiktoren 51
 Zeitabhängige Prädiktoren definieren. 51
Untergruppen 52
Modell 52
Statistik. 52
Diagramme 53
Hypothesentests. 54
Speichern 54
Export 55
Optionen 57
Zusätzliche Funktionen beim Befehl CSCOXREG . . 58

Bemerkungen. 59

Marken. 61

Index 63

Kapitel 1. Einführung in die Prozeduren von Complex Samples

Eine Grundannahme bei analytischen Prozeduren in herkömmlichen Softwarepaketen ist, dass die Beobachtungen in einer Datendatei eine einfache Zufallsstichprobe aus der zu betrachtenden Grundgesamtheit darstellen. Diese Annahme ist für eine wachsende Anzahl von Unternehmen und Wissenschaftler unhaltbar, für die es kostengünstig und zweckmäßig ist, Stichproben auf strukturiertere Weise zu gewinnen.

Mit der Option "Complex Samples" (Komplexe Stichproben) können Sie eine Stichprobe nach einem komplexen Design auswählen und die Designspezifikationen in die Datenanalyse integrieren, um sicherzustellen, dass die Ergebnisse gültig sind.

Eigenschaften komplexer Stichproben

Eine komplexe Stichprobe kann sich in verschiedener Hinsicht von einer einfachen Zufallsstichprobe unterscheiden. Bei einer einfachen Zufallsstichprobe werden die einzelnen Stichprobeneinheiten zufällig mit gleicher Wahrscheinlichkeit und ohne Zurücklegen (OZ) aus der gesamten Grundgesamtheit ausgewählt. Im Gegensatz dazu kann eine komplexe Stichprobe einige oder alle der folgenden Merkmale aufweisen:

Schichtung. Bei einer geschichteten Stichprobenziehung werden die Stichproben unabhängig voneinander innerhalb von sich nicht überschneidenden Untergruppen der Grundgesamtheit, den sogenannten Schichten, ausgewählt. Beispiele für Schichten sind sozioökonomische Gruppen, Berufsgruppen, Altersgruppen oder ethnische Gruppen. Bei Verwendung einer Schichtung können Sie angemessene Stichprobengrößen für zu untersuchende Untergruppen gewährleisten, die Genauigkeit von Gesamtschätzungen verbessern und unterschiedliche Stichprobenverfahren für die verschiedenen Schichten verwenden.

Clustering. Zur Ziehung von Clusterstichproben gehört die Auswahl von Gruppen von Stichprobeneinheiten, sogenannter Cluster. Beispiele für Cluster sind Schulen, Krankenhäuser oder geografische Gebiete; die dazugehörigen Stichprobeneinheiten sind Schüler, Patienten bzw. Einwohner. Clusterbildung ist bei mehrstufigen Plänen und Gebietsstichproben (geografischen Stichproben) üblich.

Mehrere Stufen. Bei einer mehrstufigen Stichprobenziehung wird zunächst auf der Grundlage von Clustern eine Stichprobe für die erste Stufe ausgewählt. Dann wird eine Stichprobe der zweiten Stufe ausgewählt, indem aus den ausgewählten Clustern Teilstichproben gezogen werden. Wenn die Stichprobe der zweiten Stufe auf Teilkumpen beruht, können Sie eine dritte Stufe zur Stichprobe hinzufügen. In der ersten Stufe einer Umfrage könnte beispielsweise eine Stichprobe von Städten gezogen werden. Aus den ausgewählten Städten könnten dann Stichproben der Haushalte gezogen werden. Schließlich könnten einzelne Personen aus den ausgewählten Haushalten befragt werden. Mit dem Stichproben- und dem Analyse vorbereitungsassistenten können Sie drei Stufen in einem Plan angeben.

Ziehen nicht zufälliger Stichproben. Wenn eine zufällige Auswahl schwer zu erzielen ist, können die Stichprobeneinheiten systematisch (in festgelegten Intervallen) oder sequenziell gezogen werden.

Ungleiche Auswahlwahrscheinlichkeiten. Bei der Ziehung von Clustern, die jeweils eine andere Anzahl von Einheiten enthalten, können Sie eine PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) für die Stichprobenziehung verwenden. Diese gewährleistet, dass die Auswahlwahrscheinlichkeit eines Clusters dem Anteil an Einheiten entspricht, die er enthält. Bei der PPS-Stichprobenziehung können auch allgemeinere Gewichtungsschemas für die Auswahl der Einheiten verwendet werden.

Unbeschränkte Stichprobenziehung. Bei der unbeschränkten Stichprobenziehung werden Einheiten mit Zurücklegen (MZ) ausgewählt. Eine Einheit kann also mehrmals für die Stichprobe ausgewählt werden.

Stichprobengewichtungen. Stichprobengewichtungen werden beim Ziehen komplexer Stichproben automatisch berechnet und entsprechen idealerweise der "Häufigkeit", die jede Stichprobeneinheit in der Grundgesamtheit aufweist. Daher sollte die Summe der Gewichtungen in der Stichprobe einen Schätzwert für den Umfang der Grundgesamtheit darstellen. Für die Analyseverfahren in "Komplexe Stichproben" sind Stichprobengewichtungen für die ordnungsgemäße Analyse komplexer Stichproben erforderlich. Hinweis: Diese Gewichtungen sollten ausschließlich in der Option "Komplexe Stichproben" und nicht bei anderen Analyseverfahren über die Prozedur "Fälle gewichten" verwendet werden. Bei der Prozedur "Fälle gewichten" werden die Gewichtungen als Fallreplikationen behandelt.

Verwendung der Prozeduren für komplexe Stichproben

Welche Verfahren für komplexe Stichproben für Sie infrage kommen, hängt von Ihren jeweiligen Bedürfnissen ab. Die Hauptbenutzertypen haben folgende Ziele:

- Planung und Durchführung von Studien anhand komplexer Designs, eventuell spätere Analyse der Stichprobe Das wichtigste Tool für Personen, die Studien durchführen, ist der Stichprobenassistent.
- Analysieren von Dateien mit Stichprobendaten, die zuvor anhand komplexer Designs gewonnen wurden Bevor Sie die Analyseverfahren für komplexe Stichproben nutzen können, benötigen Sie möglicherweise den Analysevorbereitungsassistenten.

Unabhängig davon, welcher Benutzertyp Sie sind, müssen Sie für die Prozeduren für komplexe Stichproben Planinformationen angeben. Diese Informationen werden zur einfacheren Wiederverwendung in einer **Plandatei** gespeichert.

Plandateien

Eine Plandatei enthält Festlegungen für komplexe Stichproben. Es gibt zwei Typen von Plandateien:

Stichprobenplan. Durch die im Stichprobenassistenten angegebenen Spezifikationen wird ein Stichprobendesign definiert, das zum Ziehen von komplexen Stichproben verwendet wird. Diese Spezifikationen sind in der Stichprobenplandatei enthalten. Eine Stichprobenplandatei enthält außerdem einen Standardanalyseplan, der für das angegebene Stichprobendesign geeignete Schätzmethoden verwendet.

Analyseplan. Diese Plandatei enthält Informationen, die bei den Analyseverfahren in "Komplexe Stichproben" benötigt werden, um die Varianzschätzungen für komplexe Stichproben ordnungsgemäß zu berechnen. Zum Plan gehören die Stichprobenstruktur, Schätzmethoden für die einzelnen Stufen und Verweise auf erforderliche Variablen, wie beispielsweise die Stichprobengewichtungen. Mit dem Analysevorbereitungsassistenten können Sie Analysepläne erstellen und bearbeiten.

Das Speichern der Angaben in einer Plandatei bringt verschiedene Vorteile mit sich, unter anderem folgende:

- Personen, die Studien durchführen, können die erste Stufe eines mehrstufigen Stichprobenplans angeben und die Einheiten der ersten Stufe sofort ziehen, Informationen für die Ziehung der Stichprobeneinheiten der zweiten Stufe sammeln und dann die zweite Stufe in den Stichprobenplan integrieren.
- Ein Analytiker, dem die Stichprobenplandatei nicht zugänglich ist, kann einen Analyseplan angeben und bei jedem Analyseverfahren für komplexe Stichproben auf diesen Plan zurückgreifen.
- Ein Entwickler großer öffentlich zugänglicher Stichproben (Public-Use-Stichproben) kann die Stichprobenplandatei veröffentlichen und damit die Anweisungen für die Analytiker vereinfachen und ermöglichen, dass nicht jeder Analytiker einen eigenen Analyseplan entwickeln muss.

Weiterführende Literatur

Weitere Informationen zu Stichprobenverfahren finden Sie in folgenden Texten:

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Kapitel 2. Stichprobenziehung mithilfe eines komplexen Designs

Der Stichprobenassistent führt Sie durch die Schritte zum Erstellen, Bearbeiten bzw. Ausführen einer Stichprobenplandatei. Vor der Verwendung des Assistenten sollten Sie über eine klar umrissene Zielgrundgesamtheit und eine Liste der Stichprobeneinheiten verfügen und ein geeignetes Stichprobendesign im Kopf haben.

Erstellen eines neuen Stichprobenplans

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Stichprobe auswählen...
2. Wählen Sie die Option **Stichprobe entwerfen** und wählen Sie einen Dateinamen für die Plandatei, in der der Stichprobenplan gespeichert werden soll.
3. Klicken Sie auf **Weiter**, um unter Verwendung des Assistenten fortzufahren.
4. Optional können Sie im Schritt "Designvariablen" Schichten, Cluster und Eingabestichprobengewichtungen definieren. Klicken Sie anschließend auf **Weiter**.
5. Optional können Sie im Schritt "Methode der Stichprobenziehung" eine Methode für die Auswahl der Items auswählen.
Bei Auswahl von **PPS Brewer** oder **PPS Murthy** können Sie auf **Fertigstellen** klicken, um die Stichprobe zu ziehen. Andernfalls klicken Sie auf **Weiter** und gehen Sie dann folgendermaßen vor:
6. Geben Sie im Schritt "Stichprobenumfang" die Anzahl bzw. den Anteil der Einheiten für die Stichprobenziehung an.
7. Jetzt können Sie auf **Fertigstellen** klicken, um die Stichprobe zu ziehen.

Optional können Sie in weiteren Schritten folgende Aktionen durchführen:

- Wählen Sie Ausgabevariablen aus, die gespeichert werden sollen.
- Fügen Sie dem Plan eine zweite oder dritte Stufe hinzu.
- Legen Sie verschiedene Auswahloptionen fest, beispielsweise die Folgenden: die Angabe, aus welchen Stufen die Stichproben gezogen werden sollen, der Startwert für Zufallszahlen und die Angabe, ob benutzerdefiniert fehlende Werte als gültige Werte von Designvariablen behandelt werden sollen.
- Wählen Sie den Speicherort für die Ausgabedaten aus.
- Fügen Sie die Auswahlen als Befehlssyntax ein.

Stichprobenassistent: Designvariablen

In diesem Schritt können Sie die Schichtungs- und Clustervariablen auswählen und Eingabestichprobengewichtungen definieren. Außerdem können Sie eine Beschriftung für die Stufe angeben.

Schichten nach. Durch die Kombination von Schichtungsvariablen werden eindeutige Teilgesamtheiten, sogenannte Schichten, definiert. Aus jeder Schicht werden getrennte Stichproben gewonnen. Zur Verbesserung der Genauigkeit Ihrer Schätzungen sollten Einheiten innerhalb von Schichten für die zu untersuchenden Merkmale so homogen wie möglich sein.

Cluster. Clustervariablen definieren Gruppen von Beobachtungseinheiten, sogenannte Cluster. Die Definition von Clustern ist sinnvoll, wenn eine unmittelbare Stichprobenziehung der Beobachtungseinheiten aus der Grundgesamtheit kostspielig oder nicht möglich ist. Stattdessen können Sie Cluster aus der Grundgesamtheit ziehen und dann aus den ausgewählten Clustern die Stichprobe der Beobachtungseinheiten ziehen. Die Verwendung von Clustern kann jedoch zu Korrelationen zwischen den Stichprobeneinheiten

führen, wodurch die Genauigkeit verringert wird. Um diese Wirkung möglichst gering zu halten, sollten die Einheiten innerhalb der Cluster für die zu untersuchenden Merkmale so heterogen wie möglich sein. Für einen mehrstufigen Plan muss mindestens eine Clustervariable definiert werden. Cluster sind außerdem erforderlich, wenn mehrere verschiedene Stichprobenmethoden verwendet werden sollen. Weitere Informationen hierzu finden Sie unter „Stichprobenassistent: Methode der Stichprobenziehung“.

Eingabestichprobengewichtung. Wenn das aktuelle Stichprobendesign Teil eines größeren Stichprobendesigns ist, können Stichprobengewichtungen aus einer früheren Stufe des größeren Designs vorliegen. In der ersten Stufe des aktuellen Plans können Sie eine numerische Variable angeben, die diese Gewichtungen enthält. Die Stichprobengewichtungen für die weiteren Stufen des aktuellen Plans werden automatisch berechnet.

Stufenbeschriftung. Sie können für jede Stufe eine optionale Zeichenfolgebeschriftung angeben. Dieses wird in der Ausgabe verwendet, um die stufenweisen Informationen besser identifizieren zu können.

Hinweis: Die Liste der Quellenvariablen hat in allen Schritten des Assistenten denselben Inhalt. Anders ausgedrückt: Variablen, die in einem Schritt aus der Liste der Quellenvariablen entfernt werden, werden in allen Schritten aus der Liste entfernt. Variablen, die wieder zur Liste der Quellenvariablen hinzugefügt werden, werden in allen Schritten in der Liste angezeigt.

Baumstruktursteuerungen zur Navigation im Stichprobenassistenten

Auf der linken Seite jedes Schritts im Stichprobenassistenten finden Sie eine Gliederung, die eine Übersicht über alle Schritte bietet. Sie können im Assistenten navigieren, indem Sie in der Gliederung auf den Namen eines aktivierten Schrittes klicken. Schritte sind aktiviert, wenn alle vorherigen Schritte gültig sind, d. h. wenn für jeden vorherigen Schritt die erforderlichen Mindestangaben vorgenommen wurden. Weitere Informationen dazu, warum ein Schritt möglicherweise ungültig ist, finden Sie in der Hilfe zu den einzelnen Schritten.

Stichprobenassistent: Methode der Stichprobenziehung

In diesem Schritt können Sie angeben, wie Fälle aus dem aktiven Dataset ausgewählt werden sollen.

Methode. Die Steuerelemente in dieser Gruppe werden zur Festlegung einer Auswahlmethode verwendet. Bei einigen Arten der Stichprobenziehung können Sie auswählen, ob die Stichprobenziehung mit Zurücklegen (MZ) oder ohne Zurücklegen (OZ) erfolgen soll. Weitere Informationen finden Sie in den Beschreibungen zu den verschiedenen Arten. Hinweis: einige PPS-Methoden (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) sind nur verfügbar, wenn Cluster definiert wurden, und alle PPS-Methoden sind nur in der ersten Stufe eines Plans verfügbar. Außerdem sind Methoden MZ nur in der letzten Stufe eines Plans verfügbar.

- **Einfache Zufallsstichprobenziehung.** Die Einheiten werden mit gleicher Wahrscheinlichkeit ausgewählt. Sie können mit oder ohne Zurücklegen ausgewählt werden.
- **Einfach systematisch.** Die Einheiten werden in festgelegten Intervallen im gesamten Stichprobenrahmen (bzw. den Schichten, wenn sie angegeben wurden) ausgewählt und ohne Zurücklegen gezogen. Eine zufällig ausgewählte Einheit im ersten Intervall wird als Startpunkt festgelegt.
- **Einfach sequenziell.** Die Einheiten werden sequenziell mit gleicher Wahrscheinlichkeit und ohne Zurücklegen ausgewählt.
- **PPS.** Dies ist eine Methode für die erste Stufe, bei der Einheiten zufällig ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zum Umfang. Alle Einheiten können mit Zurücklegen ausgewählt werden; nur Cluster können ohne Zurücklegen ausgewählt werden.
- **PPS systematisch.** Dies ist eine Methode für die erste Stufe, bei der Einheiten systematisch ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zum Umfang. Die Auswahl erfolgt ohne Zurücklegen.
- **PPS sequenziell.** Dies ist eine Methode für die erste Stufe, bei der Einheiten sequenziell ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Clustergröße.

- **PPS Brewer.** Dies ist eine Methode für die erste Stufe, bei der aus jeder Schicht zwei Cluster ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Clustergröße. Damit diese Methode verwendet werden kann, muss eine Clustervariable angegeben werden.
- **PPS Murthy.** Dies ist eine Methode für die erste Stufe, bei der aus jeder Schicht zwei Cluster ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Clustergröße. Damit diese Methode verwendet werden kann, muss eine Clustervariable angegeben werden.
- **PPS Sampford.** Dies ist eine Methode für die erste Stufe, bei der aus jeder Schicht mehr als zwei Cluster ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Clustergröße. Es handelt sich um eine Erweiterung der Brewer-Methode. Damit diese Methode verwendet werden kann, muss eine Clustervariable angegeben werden.
- **MZ-Schätzung für Analyse verwenden.** Standardmäßig wird in der Plandatei eine Schätzmethode angegeben, die mit der ausgewählten Stichprobenmethode konsistent ist. Dadurch können Sie eine Schätzung mit Zurücklegen verwenden, selbst wenn die Stichprobenmethode eine Schätzung ohne Zurücklegen beinhaltet. Diese Option ist nur in Stufe 1 verfügbar.

Maß für die Größe der Einheiten (MOS). Bei Auswahl einer PPS-Methode müssen Sie ein Maß für die Größe angeben, mit dem die Größe jeder Einheit festgelegt wird. Diese Größen können explizit in einer Variablen definiert oder aus den Daten berechnet werden. Optional können Sie für das MOS Unter- und Obergrenzen festlegen, die Vorrang vor allen Werten haben, die in der MOS-Variablen gefunden oder aus den Daten berechnet werden. Diese Optionen sind nur in Stufe 1 verfügbar.

Stichprobenassistent: Stichprobenumfang

In diesem Schritt können Sie die Anzahl bzw. den Anteil der Einheiten festlegen, die in der aktuellen Stufe als Stichprobe gezogen werden sollen. Der Stichprobenumfang kann fest oder für die verschiedenen Schichten unterschiedlich sein. Für die Angabe des Stichprobenumfangs können die in den vorherigen Stufen ausgewählten Clustern verwendet werden, um Schichten zu definieren.

Einheiten. Sie können einen genauen Stichprobenumfang oder den Anteil der Einheiten für die Stichprobe angeben.

- **Wert.** Allen Schichten wird derselbe Wert zugewiesen. Wenn **Anzahl** als Metrik für die Einheiten ausgewählt wurde, sollten Sie eine positive Ganzzahl eingeben. Bei Auswahl von **Anteile** sollten Sie einen nicht negativen Wert eingeben. Außer bei Stichprobenziehung mit Zurücklegen dürfen die Anteilswerte außerdem nicht größer als 1 sein.
- **Ungleiche Werte für Schichten.** Ermöglicht die Eingabe von Umfangswerten für die einzelnen Schichten über das Dialogfeld "Ungleiche Werte für Schichten".
- **Werte aus Variable lesen.** Ermöglicht die Auswahl einer numerischen Variablen, die die Umfangswerte für Schichten enthält.

Bei Auswahl von **Anteile** haben Sie die Möglichkeit, Unter- und Obergrenzen für die Anzahl der in der Stichprobe enthaltenen Einheiten festzulegen.

Ungleiche Umfänge definieren

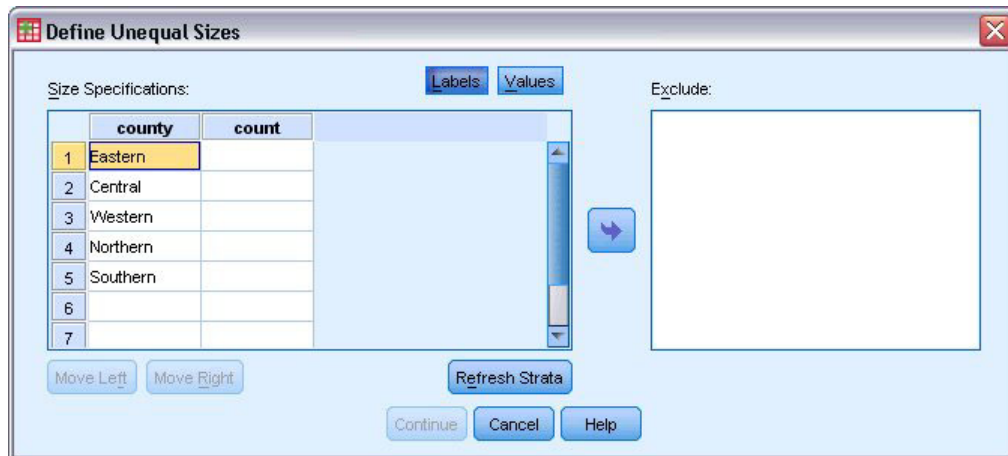


Abbildung 1. Dialogfeld "Ungleiche Umfänge definieren"

Im Dialogfeld "Ungleiche Umfänge definieren" können Sie Umfänge schichtweise eingeben.

Raster "Spezifikationen für Umfang". Das Raster zeigt die Kombination von bis zu fünf Schicht- oder Clustervariablen an, wobei Zeile eine Schicht-/Clusterkombination enthält. Als Rastervariablen zulässig sind alle Schichtungsvariablen aus der aktuellen Stufe und den vorherigen Stufen sowie alle Clustervariablen aus den vorherigen Stufen. Die Variablen können im Raster neu angeordnet oder in die Ausschlussliste verschoben werden. Geben Sie die Werte für den Umfang in die Spalte rechts außen ein. Klicken Sie auf **Beschriftungen** oder **Werte**, um die Anzeige der Wertbeschriftungen und der Datenwerte für die Schichtungs- und Clustervariablen in den Rasterzellen ein- bzw. auszuschalten. Bei Zellen, die Werte ohne Beschriftungen enthalten, werden immer Werte angezeigt. Klicken Sie auf **Schichten aktualisieren**, um das Raster mit allen Kombinationen von beschrifteten Datenwerten für Variablen im Raster neu auszufüllen.

Ausschließen. Um die Umfänge für ein Subset von Schicht-/Clusterkombinationen anzugeben, verschieben Sie eine oder mehrere Variablen in die Ausschlussliste. Diese Variablen werden nicht für die Festlegung der Stichprobenumfänge verwendet.

Stichprobenassistent: Ausgabevariablen

In diesem Schritt können Sie Variablen auswählen, die beim Ziehen der Stichprobe gespeichert werden sollen.

Umfang der Grundgesamtheit. Die geschätzte Anzahl an Einheiten in der Grundgesamtheit für eine bestimmte Stufe. Der Stammname der gespeicherten Variablen lautet *PopulationSize_*.

Stichprobenanteil. Die Stichprobenquote in einer bestimmten Stufe. Der Stammname der gespeicherten Variablen lautet *SamplingRate_*.

Umfang der Stichprobe. Die Anzahl der Einheiten, die in einer bestimmten Stufe als Stichprobe gezogen wurden. Der Stammname der gespeicherten Variablen lautet *SampleSize_*.

Stichprobengewichtung. Dies ist die Inverse der Einschlusswahrscheinlichkeiten. Der Stammname der gespeicherten Variablen lautet *SampleWeight_*.

Einige stufenweise Variablen werden automatisch generiert. Dazu gehören:

Einschlusswahrscheinlichkeiten. Der Anteil der Einheiten, die in einer bestimmten Stufe als Stichprobe gezogen wurden. Der Stammname der gespeicherten Variablen lautet *InclusionProbability_*.

Kumulative Gewichtung. Die kumulative Stichprobengewichtung über alle Stufen einschließlich der aktuellen. Der Stammname der gespeicherten Variablen lautet *SampleWeightCumulative_*.

Index. Identifiziert Einheiten, die mehrmals in einer Stufe ausgewählt wurden. Der Stammname der gespeicherten Variablen lautet *Index_*.

Hinweis: Die Stammmamen der gespeicherten Variablen beinhalten ein ganzzahliges Suffix, das der Stufennummer entspricht, beispielsweise *PopulationSize_1_* für die gespeicherte Größe des Stichprobenumfangs für Stufe 1.

Stichprobenassistent: Planübersicht

Dies ist der letzte Schritt in jeder Stufe, in der eine Übersicht der in der aktuellen Stufe vorgenommenen Angaben zum Stichprobendesign angezeigt wird. Anschließend können Sie entweder zur nächsten Stufe weitergehen (und sie, falls erforderlich, erstellen) oder die Optionen für die Stichprobenziehung festlegen.

Stichprobenassistent: Stichprobe ziehen: Auswahloptionen

In diesem Schritt können Sie auswählen, ob eine Stichprobe gezogen werden soll. Außerdem können Sie andere Optionen für die Stichprobenauswahl festlegen, beispielsweise den Startwert für Zufallszahlen und den Umgang mit fehlenden Werten.

Stichprobe ziehen. Sie können auswählen, ob eine Stichprobe gezogen werden soll. Außerdem können Sie festlegen, dass ein Teil des Stichprobendesigns ausgeführt werden soll. Die Stichprobenziehungen für die Stufen muss der Reihe nach erfolgen. Die Stichprobenziehung für Stufe 2 kann also erst erfolgen, wenn die Stichprobe für Stufe 1 gezogen wurde. Bei der Bearbeitung oder Ausführung eines Plans kann für gesperrte Stufen keine erneute Stichprobenziehung erfolgen.

Startwert. Damit können Sie einen Startwert für die Generierung von Zufallszahlen auswählen.

Benutzerdefiniert fehlende Werte einschließen. Hiermit wird festgelegt, ob benutzerdefiniert fehlende Werte gültig sind. Wenn ja, werden benutzerdefiniert fehlende Werte als gesonderte Kategorie behandelt.

Daten sind schon sortiert. Wenn der Stichprobenrahmen anhand der Werte der Schichtungsvariablen vortriert wurde, können Sie mit dieser Option den Auswahlvorgang beschleunigen.

Stichprobenassistent: Stichprobe ziehen: Ausgabedateien

In diesem Schritt können Sie die Ausgabedateien für die in der Stichprobe enthaltenen Fälle, die Gewichtsvariablen, gemeinsamen Wahrscheinlichkeiten und Fallauswahlregeln angeben.

Stichprobendaten. Mit diesen Optionen können Sie festlegen, wo die Ausgabedaten für die Stichprobe gespeichert werden sollen. Sie können dem aktiven Dataset hinzugefügt, in ein neues Dataset geschrieben oder in einer externen Datendatei im IBM® SPSS Statistics-Format gespeichert werden. Datasets sind in der aktuellen Sitzung verfügbar. In nachfolgenden Sitzungen stehen Datasets nur zur Verfügung, wenn Sie sie ausdrücklich als Datendateien speichern. Die Namen von Datasets müssen den Regeln zum Benennen von Variablen entsprechen. Wenn eine externe Datei oder ein neues Dataset angegeben wird, werden die Ausgabevariablen für die Stichprobenziehung und die Variablen im aktiven Dataset für die ausgewählten Fälle gespeichert.

Gemeinsame Wahrscheinlichkeiten. Mit diesen Optionen können Sie festlegen, in welche Datei die gemeinsamen Wahrscheinlichkeiten geschrieben werden sollen. Sie werden in einer externen Datendatei im IBM SPSS Statistics-Format gespeichert. Gemeinsame Wahrscheinlichkeiten werden erstellt, wenn die PPS-Methode ohne Zurücklegen, "PPS Brewer", "PPS Sampford" oder "PPS Murthy" ausgewählt und "MZ-Schätzung" nicht angegeben ist.

Fallauswahlregeln. Wenn Sie die Stichprobe für jede Stufe einzeln erstellen, sollten Sie die Fallauswahlregeln in einer Textdatei speichern. Die Fallauswahlregeln unterstützen die Erstellung des Teilrahmens für die weiteren Stufen.

Stichprobenassistent: Fertigstellen

Dies ist der letzte Schritt. Sie können die Plandatei jetzt speichern und die Stichprobe ziehen oder Ihre Auswahl in ein Syntaxfenster einfügen.

Wenn Sie Änderungen an Schritten in der bestehenden Plandatei vornehmen, können Sie den bearbeiteten Plan in einer neuen Datei speichern oder die bestehende Plandatei überschreiben. Wenn Sie Stufen hinzufügen, ohne Änderungen an bestehenden Stufen vorzunehmen, überschreibt der Assistent die bestehende Plandatei. Wenn Sie den Plan in einer neuen Datei speichern möchten, wählen Sie die Option **Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen** aus und ändern Sie den Dateinamen in den Syntaxbefehlen.

Bearbeiten eines bestehenden Stichprobenplans

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Stichprobe auswählen...
2. Wählen Sie die Option **Stichprobendesign bearbeiten** und wählen Sie eine Plandatei zur Bearbeitung aus.
3. Klicken Sie auf **Weiter**, um unter Verwendung des Assistenten fortzufahren.
4. Überarbeiten Sie den Stichprobenplan im Schritt "Planübersicht" und klicken Sie auf **Weiter**.
Die darauf folgenden Schritte sind größtenteils mit denen für einen neuen Plan identisch. Weitere Informationen finden Sie in der Hilfe zu den einzelnen Schritten.
5. Wechseln Sie zum Schritt "Fertigstellen" und geben Sie einen neuen Namen für die bearbeitete Plandatei an oder legen Sie fest, dass die bestehende Plandatei überschrieben werden soll.

Die folgenden Optionen sind verfügbar:

- Geben Sie Stufen an, für die die Stichprobenziehung bereits erfolgt ist.
- Entfernen Sie Stufen aus dem Plan.

Stichprobenassistent: Planübersicht

In diesem Schritt können Sie den Stichprobenplan überprüfen und Stufen angeben, für die die Stichprobenziehung bereits erfolgt ist. Wenn Sie einen Plan bearbeiten, können Sie auch Stufen aus dem Plan entfernen.

Stufen mit erfolgter Stichprobenziehung. Wenn kein erweiterter Stichprobenrahmen verfügbar ist, müssen Sie ein mehrstufiges Stichprobendesign für jede Stufe einzeln ausführen. Wählen Sie in der Dropdown-Liste aus, für welche Stufen die Stichprobenziehung bereits erfolgt ist. Alle Stufen, die bereits ausgeführt wurden, sind gesperrt, stehen also im Schritt "Stichprobe ziehen: Auswahloptionen" nicht zur Verfügung und können beim Bearbeiten des Plans nicht geändert werden.

Stufen entfernen. Sie können die Stufen 2 und 3 aus einem mehrstufigen Plan entfernen.

Ausführen eines bestehenden Stichprobenplans

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Stichprobe auswählen...
2. Wählen Sie die Option **Stichprobe ziehen** und wählen Sie eine Plandatei für die Ausführung aus.
3. Klicken Sie auf **Weiter**, um unter Verwendung des Assistenten fortzufahren.
4. Überarbeiten Sie den Stichprobenplan im Schritt "Planübersicht" und klicken Sie auf **Weiter**.
5. Die einzelnen Schritte mit Informationen zur Stufe werden bei der Ausführung eines Stichprobenplans übersprungen. Sie können nun jederzeit mit dem Schritt "Fertigstellen" fortfahren.

Optional können Sie die Stufen angeben, für die die Stichprobenziehung bereits erfolgt ist.

Zusätzliche Funktionen bei den Befehlen CSPLAN und CSSELECT

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Legen Sie die benutzerdefinierten Namen für Ausgabevariablen fest.
- Legen Sie die Ausgabe im Viewer fest. Sie können beispielsweise die stufenweise Übersicht über den Plan unterdrücken, der angezeigt wird, wenn eine Stichprobe entworfen oder bearbeitet wird, die Übersicht über die Fälle in der Stichprobe für die einzelnen Schichten unterdrücken und eine Zusammenfassung der Fallverarbeitung abrufen.
- Wählen Sie ein Subset der Variablen im aktiven Dataset aus, die in eine externe Stichprobendatei oder in ein anderes Dataset geschrieben werden soll.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 3. Vorbereiten einer komplexen Stichprobe für die Analyse

Der Analysevorbereitungsassistent führt Sie durch die Schritte zum Erstellen bzw. Bearbeiten eines Analyseplans zur Verwendung mit den verschiedenen Analyseverfahren für komplexe Stichproben. Vor der Verwendung des Assistenten sollten Sie nach einem komplexen Design eine Stichprobe gezogen haben.

Das Erstellen eines neuen Plans ist am sinnvollsten, wenn Sie keinen Zugriff auf die Datei mit dem Stichprobenplan haben, der zum Ziehen der Stichprobe verwendet wurde (der Stichprobenplan enthält einen Standardanalyseplan). Wenn Sie Zugriff auf die Datei mit dem Stichprobenplan haben, der zum Ziehen der Stichprobe verwendet wurde, können Sie den in der Datei enthaltenen Standardanalyseplan verwenden oder die Standardfestlegungen für die Analyse abändern und Ihre Änderungen in einer neuen Datei speichern.

Erstellen eines neuen Analyseplans

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Für Analyse vorbereiten...
2. Wählen Sie die Option **Plandatei erstellen** aus und wählen Sie einen Dateinamen für die Plandatei, in der der Analyseplan gespeichert werden soll.
3. Klicken Sie auf **Weiter**, um unter Verwendung des Assistenten fortzufahren.
4. Geben Sie die Variable mit den Stichprobengewichten im Schritt "Designvariablen" an. Definieren Sie gegebenenfalls Schichten und Cluster.
5. Jetzt können Sie auf **Fertigstellen** klicken, um den Plan zu speichern.

Optional können Sie in weiteren Schritten folgende Aktionen durchführen:

- Wählen Sie die Methode zum Schätzen der Standardfehler im Schritt "Schätzmethode" aus.
- Geben Sie die Anzahl der Einheiten in der Stichprobe oder der Einschlusswahrscheinlichkeit pro Einheit im Schritt "Umfang" an.
- Fügen Sie dem Plan eine zweite oder dritte Stufe hinzu.
- Fügen Sie die Auswahlen als Befehlssyntax ein.

Analysevorbereitungsassistent: Designvariablen

In diesem Schritt können Sie die Schichtungs- und Clustervariablen identifizieren und Stichprobengewichtungen definieren. Außerdem können Sie eine Beschriftung für die Stufe angeben.

Schichten. Durch die Kombination von Schichtungsvariablen werden eindeutige Teilgesamtheiten, sogenannte Schichten, definiert. Die gesamte Stichprobe besteht aus der Kombination unabhängiger Stichproben aus jeder Schicht.

Cluster. Clustervariablen definieren Gruppen von Beobachtungseinheiten, sogenannte Cluster. Bei in mehreren Stufen gezogenen Stichproben werden Cluster in den früheren Stufen und anschließend Teilstichproben aus den ausgewählten Clustern gezogen. Beim Analysieren von Datendateien, die durch die Ziehung von Clusterstichproben mit Zurücklegen erstellt werden, sollten Sie auch den Duplizierungsindex als Clustervariable verwenden.

Stichprobengewichtung. In der ersten Stufe müssen Sie Stichprobengewichtungen angeben. Die Stichprobengewichtungen für die weiteren Stufen des aktuellen Plans werden automatisch berechnet.

Stufenbeschriftung. Sie können für jede Stufe eine optionale Zeichenfolgebeschriftung angeben. Dieses wird in der Ausgabe verwendet, um die stufenweisen Informationen besser identifizieren zu können.

Hinweis: Die Liste der Quellenvariablen hat in allen Schritten des Assistenten denselben Inhalt. Anders ausgedrückt: Variablen, die in einem Schritt aus der Liste der Quellenvariablen entfernt werden, werden in allen Schritten aus der Liste entfernt. Variablen, die wieder zur Liste der Quellenvariablen hinzugefügt werden, erscheinen in allen Schritten.

Baumstruktursteuerungen zur Navigation im Analyseassistenten

Auf der linken Seite jedes Schritts im Analyseassistenten finden Sie eine Gliederung, die eine Übersicht über alle Schritte bietet. Sie können im Assistenten navigieren, indem Sie in der Gliederung auf den Namen eines aktivierten Schrittes klicken. Schritte sind aktiviert, wenn alle vorherigen Schritte gültig sind d. h., solange für jeden vorherigen Schritt die erforderlichen Mindestangaben vorgenommen wurden. Weitere Informationen dazu, warum ein Schritt möglicherweise ungültig ist, finden Sie in der Hilfe zu den einzelnen Schritten.

Analysevorbereitungsassistent: Schätzmethode

In diesem Schritt können Sie eine Schätzmethode für die Stufe angeben.

MZ (Stichprobenziehung mit Zurücklegen). Die MZ-Schätzung beinhaltet keine Endlichkeitskorrektur, (Finite Population Correction, FPC) bei der Schätzung der Varianz in einem komplexen Stichprobendesign. Beim Schätzen der Varianz bei einer einfachen Zufallsstrichprobenziehung (Simple Random Sampling, SRS) können Sie auswählen, ob die Endlichkeitskorrektur (FPC) aufgenommen oder ausgeschlossen werden soll.

Es wird empfohlen, bei der SRS-Varianzschätzung keine FPC aufzunehmen, wenn die Analysegewichtungen skaliert wurden, sodass ihre Summe nicht den Umfang der Grundgesamtheit ergibt. Die SRS-Varianzschätzung wird bei der Berechnung von Statistiken wie dem Designeffekt verwendet. MZ-Schätzung kann nur in der letzten Stufe eines Planes angegeben werden; der Assistent lässt nicht zu, dass eine weitere Stufe hinzugefügt wird, wenn die Option für die MZ-Schätzung ausgewählt wird.

OZ gleich (Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen). Die Schätzung für "OZ gleich" beinhaltet eine Endlichkeitskorrektur und geht davon aus, dass die Stichprobenziehung bei den Einheiten mit gleicher Wahrscheinlichkeit erfolgt. "OZ gleich" kann in jeder Stufe eines Plans angegeben werden.

OZ ungleich (Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen). Neben der Endlichkeitskorrektur berücksichtigt "OZ ungleich" auch Stichprobeneinheiten (in der Regel Cluster), die mit ungleicher Wahrscheinlichkeit ausgewählt wurden. Diese Schätzmethode ist nur in der ersten Stufe verfügbar.

Analysevorbereitungsassistent: Umfang

Dieser Schritt dient zur Angabe der Einschlusswahrscheinlichkeiten bzw. der Umfänge der Grundgesamtheiten für die aktuelle Stufe. Die Umfänge können fest oder für die verschiedenen Schichten unterschiedlich sein. Für die Angabe der Umfänge können die in den vorherigen Stufen festgelegten Cluster verwendet werden, um Schichten zu definieren. Beachten Sie, dass dieser Schritt nur erforderlich ist, wenn als Schätzmethode "OZ gleich" ausgewählt wurde.

Einheiten. Sie können den genauen Umfang der Grundgesamtheiten angeben oder die Wahrscheinlichkeiten, mit denen die Stichprobenziehung der Einheiten erfolgte.

- **Wert.** Allen Schichten wird derselbe Wert zugewiesen. Wenn **Umfang der Grundgesamtheiten** als Metrik für die Einheiten ausgewählt wurde, sollten Sie eine nicht negative Ganzzahl eingeben. Bei Auswahl von **Einschlusswahrscheinlichkeiten** sollten Sie einen Wert aus dem Bereich von 0 bis 1 eingeben.
- **Ungleiche Werte für Schichten.** Ermöglicht die Eingabe von Umfangswerten für die einzelnen Schichten über das Dialogfeld "Ungleiche Werte für Schichten".
- **Werte aus Variable lesen.** Ermöglicht die Auswahl einer numerischen Variablen, die die Umfangswerte für Schichten enthält.

Ungleiche Umfänge definieren

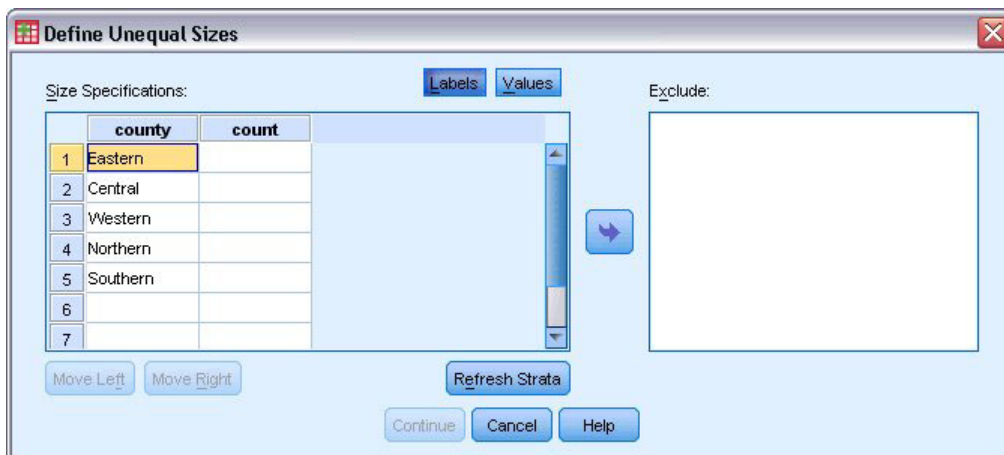


Abbildung 2. Dialogfeld "Ungleiche Umfänge definieren"

Im Dialogfeld "Ungleiche Umfänge definieren" können Sie Umfänge schichtweise eingeben.

Raster "Spezifikationen für Umfang". Das Raster zeigt die Kombination von bis zu fünf Schicht- oder Clustervariablen an, wobei Zeile eine Schicht-/Clusterkombination enthält. Als Rastervariablen zulässig sind alle Schichtungsvariablen aus der aktuellen Stufe und den vorherigen Stufen sowie alle Clustervariablen aus den vorherigen Stufen. Die Variablen können im Raster neu angeordnet oder in die Ausschlussliste verschoben werden. Geben Sie die Werte für den Umfang in die Spalte rechts außen ein. Klicken Sie auf **Beschriftungen** oder **Werte**, um die Anzeige der Wertbeschriftungen und der Datenwerte für die Schichtungs- und Clustervariablen in den Rasterzellen ein- bzw. auszuschalten. Bei Zellen, die Werte ohne Beschriftungen enthalten, werden immer Werte angezeigt. Klicken Sie auf **Schichten aktualisieren**, um das Raster mit allen Kombinationen von beschrifteten Datenwerten für Variablen im Raster neu auszufüllen.

Ausschließen. Um die Umfänge für ein Subset von Schicht-/Clusterkombinationen anzugeben, verschieben Sie eine oder mehrere Variablen in die Ausschlussliste. Diese Variablen werden nicht für die Festlegung der Stichprobenumfänge verwendet.

Analysevorbereitungsassistent: Planübersicht

Dies ist der letzte Schritt in jeder Stufe, in dem eine Übersicht der in der aktuellen Stufe vorgenommenen Angaben zum Analyseplan angezeigt wird. Anschließend können Sie entweder zur nächsten Stufe weitergehen (und sie, falls erforderlich, erstellen) oder die Analyseangaben speichern.

Wenn keine weitere Stufe hinzugefügt werden kann, hat dies vermutlich einen der folgenden Gründe:

- Im Schritt "Designvariablen" wurde keine Clustervariable angegeben.
- Im Schritt "Schätzmethode" wurde "MZ-Schätzung" angegeben.
- Dies ist die dritte Stufe der Analyse und der Assistent unterstützt maximal drei Stufen.

Analysevorbereitungsassistent: Fertigstellen

Dies ist der letzte Schritt. Sie können die Plandatei jetzt speichern oder Ihre Auswahl in ein Syntaxfenster einfügen.

Wenn Sie Änderungen an Schritten in der bestehenden Plandatei vornehmen, können Sie den bearbeiteten Plan in einer neuen Datei speichern oder die bestehende Plandatei überschreiben. Wenn Sie Stufen hinzufügen, ohne Änderungen an bestehenden Stufen vorzunehmen, überschreibt der Assistent die bestehende Plandatei. Wenn Sie den Plan in einer neuen Datei speichern möchten, wählen Sie die Option **Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen** aus und ändern Sie den Dateinamen in den Syntaxbefehlen.

Bearbeiten eines bestehenden Analyseplans

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Für Analyse vorbereiten...
2. Wählen Sie die Option **Plandatei bearbeiten** aus und wählen Sie einen Dateinamen für die Plandatei, in der der Analyseplan gespeichert werden soll.
3. Klicken Sie auf **Weiter**, um unter Verwendung des Assistenten fortzufahren.
4. Überarbeiten Sie den Analyseplan im Schritt "Planübersicht" und klicken Sie auf **Weiter**.
Die darauf folgenden Schritte sind größtenteils mit denen für einen neuen Plan identisch. Weitere Informationen finden Sie in der Hilfe zu den einzelnen Schritten.
5. Wechseln Sie zum Schritt "Fertigstellen" und geben Sie einen neuen Namen für die bearbeitete Plandatei an oder legen Sie fest, dass die bestehende Plandatei überschrieben werden soll.

Optional können Sie Stufen aus dem Plan entfernen.

Analysevorbereitungsassistent: Planübersicht

In diesem Schritt können Sie den Analyseplan überprüfen und Stufen aus dem Plan entfernen.

Stufen entfernen. Sie können die Stufen 2 und 3 aus einem mehrstufigen Plan entfernen. Da ein Plan mindestens eine Stufe aufweisen muss, können Sie die Stufe 1 zwar bearbeiten, nicht jedoch aus dem Plan entfernen.

Kapitel 4. Plan für komplexe Stichproben

Für die Analyseverfahren für komplexe Stichproben sind Analysespezifikationen aus einer Analyse- oder Stichprobenplandatei erforderlich, um gültige Ergebnisse zu erzielen.

Plan. Geben Sie den Pfad zu einer Analyse- oder Beispielplandatei an.

Gemeinsame Wahrscheinlichkeiten. Um die Schätzung "OZ ungleich" für Cluster verwenden zu können, die mit einer PPS-Methode OZ gezogen wurden, müssen Sie eine gesonderte Datei oder ein geöffnetes Dataset angeben, die bzw. das die gemeinsamen Wahrscheinlichkeiten enthält. Diese Datei bzw. dieses Dataset wird vom Stichprobenassistenten während der Stichprobenziehung erstellt.

Kapitel 5. Häufigkeiten für komplexe Stichproben

Mit der Prozedur "Häufigkeiten für komplexe Stichproben" können Sie Häufigkeitstabellen für ausgewählte Variablen erstellen und univariate Statistiken anzeigen. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur "Häufigkeiten für komplexe Stichproben" können Sie univariate Statistiken in Tabellenform für die Einnahme von Vitaminpräparaten bei US-Bürgern erstellen, die auf den Ergebnissen der Umfrage National Health Interview Survey (NHIS) beruhen und einen geeigneten Analyseplan für diese öffentlich zugänglichen Daten beinhalten.

Statistiken. Mit diesem Verfahren erhalten Sie Schätzungen für die Umfänge der Grundgesamtheiten für die Zellen und Tabellenprozentsätze, außerdem Standardfehler, Konfidenzintervalle, Variationskoeffizienten, Effekte des Stichprobenplans, Quadratwurzeln aus den Effekten des Stichprobenplans, kumulative Werte sowie die ungewichtete Anzahl für jede Schätzung. Des Weiteren werden die Chi-Quadrat-Statistik und die Likelihood-Quotienten-Statistik für den Test auf gleiche Spaltenanteile berechnet.

Erläuterungen der Daten für die Häufigkeiten für komplexe Stichproben

Daten. Variablen, für die Häufigkeitstabellen erstellt werden, sollten kategorial sein. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Berechnen von Häufigkeiten für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren > Komplexe Stichproben > Häufigkeiten...

2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.

3. Klicken Sie auf **Weiter**.

4. Wählen Sie im Dialogfeld "Häufigkeiten für komplexe Stichproben" mindestens eine Häufigkeitsvariable aus.

Optional können Sie Variablen zur Definition von Teilgesamtheiten angeben. Die Statistiken werden für jede Teilgesamtheit getrennt berechnet.

Häufigkeiten für komplexe Stichproben: Statistiken

Zellen. In dieser Gruppe können Sie Schätzungen für die Umfänge der Grundgesamtheiten für die Zellen und Tabellenprozentsätze abrufen.

Statistiken. Mit dieser Gruppe können Statistiken zum Umfang der Grundgesamtheit oder zum Tabellenprozentsatz erstellt werden.

- **Standardfehler.** Der Standardfehler der Schätzung.
- **Konfidenzintervall.** Ein Konfidenzintervall für die Schätzung unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers der Schätzung zur Schätzung.

- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung der Schätzung verwendet wurden.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Kumulative Werte.** Die kumulative Schätzung über jeden einzelnen Wert der Variablen.

Test auf gleiche Spaltenanteile. Mit dieser Option werden Chi-Quadrat- und Likelihood-Quotienten-Tests für die Hypothese erstellt, dass die Kategorien einer Variable die gleichen Häufigkeiten aufweisen. Für jede Variable werden separate Tests durchgeführt.

Komplexe Stichproben: Fehlende Werte

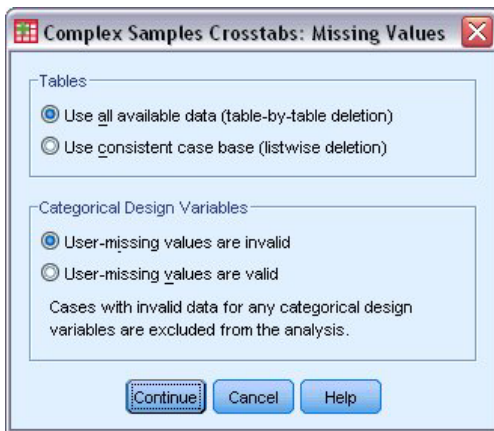


Abbildung 3. Dialogfeld "Fehlende Werte"

Tabellen. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Fehlende Werte werden jeweils für einzelne Tabellen bestimmt. Daher können die zur Berechnung von Statistiken verwendeten Fälle zwischen den einzelnen Häufigkeits- bzw. Kreuztabellen variieren.
- **Einheitliche Fallbasis verwenden.** Fehlende Werte werden über alle Variablen bestimmt. Daher sind die zur Berechnung der Statistiken verwendeten Fälle über die verschiedenen Tabellen konsistent.

Kategoriale Designvariablen. Diese Gruppe bestimmt, ob benutzerdefiniert fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

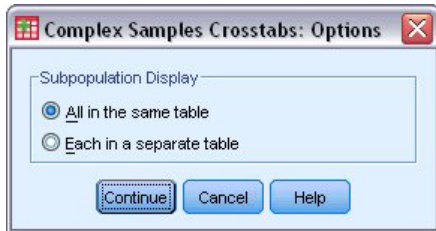


Abbildung 4. Dialogfeld "Optionen"

Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel 6. Deskriptive Statistiken für komplexe Stichproben

Die Prozedur "Deskriptive Statistiken für komplexe Stichproben" zeigt univariate Auswertungsstatistiken für verschiedene Variablen an. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur "Deskriptive Statistiken für komplexe Stichproben" können Sie univariate deskriptive Statistiken für das Aktivitätsniveau von US-Bürgern erstellen, die auf den Ergebnissen der Umfrage National Health Interview Survey (NHIS) beruhen und einen geeigneten Analyseplan für diese öffentlich zugänglichen Daten beinhalten.

Statistiken. Mit diesem Verfahren erhalten Sie Mittelwerte und Summen sowie *T*-Tests, Standardfehler, Konfidenzintervalle, Variationskoeffizienten, die ungewichteten Anzahlen, den Umfang der Grundgesamtheiten, die Effekte des Stichprobenplans und die Quadratwurzeln aus den Effekten des Stichprobenplans für jede Schätzung.

Erläuterungen der Daten für deskriptive Statistiken für komplexe Stichproben

Daten. Die Maße sollten metrische Variablen sein. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Erstellen von deskriptiven Statistiken für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren > Komplexe Stichproben > Deskriptive Statistiken...

2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.

3. Klicken Sie auf **Weiter**.

4. Wählen Sie im Dialogfeld "Deskriptive Statistiken für komplexe Stichproben" mindestens eine Maßvariable aus.

Optional können Sie Variablen zur Definition von Teilgesamtheiten angeben. Die Statistiken werden für jede Teilgesamtheit getrennt berechnet.

Deskriptive Statistiken für komplexe Stichproben: Statistiken

Auswertungen. Mit dieser Gruppe können Sie Schätzungen der Mittelwerte und Summen der Maßvariablen abrufen. Außerdem können Sie *T*-Tests der Schätzungen unter Verwendung eines angegebenen Werts abrufen.

Statistiken. Mit dieser Gruppe erhalten Sie Statistiken zu Mittelwert oder Summe.

- **Standardfehler.** Der Standardfehler der Schätzung.
- **Konfidenzintervall.** Ein Konfidenzintervall für die Schätzung unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers der Schätzung zur Schätzung.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung der Schätzung verwendet wurden.

- **Umfang der Grundgesamtheit.** Die geschätzte Anzahl an Einheiten in der Grundgesamtheit.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.

Deskriptive Statistiken für komplexe Stichproben: Fehlende Werte

Statistiken für Maßvariablen. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Die fehlenden Werte werden variablenweise bestimmt, sodass die für die Berechnung der Statistiken verwendeten Werte bei den verschiedenen Maßvariablen voneinander abweichen können.
- **Einheitliche Fallbasis gewährleisten.** Die fehlenden Werte werden über alle Variablen bestimmt, sodass die für die Berechnung der Statistiken verwendeten Fälle konsistent sind.

Kategoriale Designvariablen. Diese Gruppe bestimmt, ob benutzerdefiniert fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

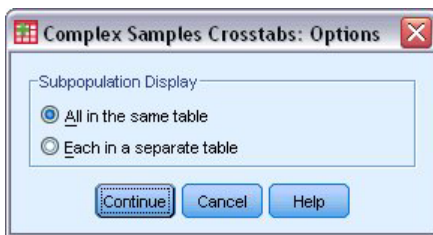


Abbildung 5. Dialogfeld "Optionen"

Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel 7. Kreuztabellen für komplexe Stichproben

Mit der Prozedur "Kreuztabellen für komplexe Stichproben" werden Kreuztabellen für Paare von ausgewählten Variablen erstellt und bivariate Statistiken angezeigt. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur "Kreuztabellen für komplexe Stichproben" können Sie Kreuzklassifikationsstatistiken für die Häufigkeit des Rauchens im Verhältnis zur Einnahme von Vitaminpräparaten bei US-Bürgern erstellen, die auf den Ergebnissen der Umfrage National Health Interview Survey (NHIS) beruhen und einen geeigneten Analyseplan für diese öffentlich zugänglichen Daten beinhalten.

Statistiken. Mit diesem Verfahren erhalten Sie Schätzungen für die Umfänge der Grundgesamtheiten der Zellen sowie Prozentsätze für Zeilen, Spalten und die Tabelle, außerdem Standardfehler, Konfidenzintervalle, Variationskoeffizienten, erwartete Werte, Effekte des Stichprobenplans, Quadratwurzeln aus den Effekten des Stichprobenplans, Residuen, korrigierte Residuen sowie die ungewichtete Anzahl für jede Schätzung. Das Odds-Verhältnis, das relative Risiko und die Risikodifferenz werden für 2x2-Tabellen berechnet. Des Weiteren werden die Pearson-Statistik und die Statistik für den Likelihood-Quotienten für den Test auf Unabhängigkeit der Zeilen- und Spaltenvariablen berechnet.

Erläuterungen der Daten für die Kreuztabellen für komplexe Stichproben

Daten. Die Zeilen- und Spaltenvariablen sollten kategorial sein. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Erstellen von Kreuztabellen für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Kreuztabellen...
2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
3. Klicken Sie auf **Weiter**.
4. Wählen Sie im Dialogfeld "Kreuztabellen für komplexe Stichproben" mindestens eine Zeilenvariable und eine Spaltenvariable aus.

Optional können Sie Variablen zur Definition von Teilgesamtheiten angeben. Die Statistiken werden für jede Teilgesamtheit getrennt berechnet.

Kreuztabellen für komplexe Stichproben - Statistik

Zellen. In dieser Gruppe können Sie Schätzungen für den Umfang der Grundgesamtheit für die Zellen sowie Prozentsätze für Zeilen, Spalten und die Tabelle abrufen.

Statistiken. In dieser Gruppe finden Sie Statistiken zu den Umfängen der Grundgesamtheiten und zu den Prozentsätzen für Zeilen, Spalten und die Tabelle.

- **Standardfehler.** Der Standardfehler der Schätzung.
- **Konfidenzintervall.** Ein Konfidenzintervall für die Schätzung unter Verwendung der festgelegten Niveaus.

- **Variationskoeffizient.** Das Verhältnis des Standardfehlers der Schätzung zur Schätzung.
- **Erwartete Werte.** Der erwartete Wert der Schätzung, unter der Hypothese der Unabhängigkeit der Zeilen- und Spaltenvariable.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung der Schätzung verwendet wurden.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Residuen.** Der erwartete Wert ist die Anzahl von Fällen, die in einer Zelle erwartet würden, wenn kein Zusammenhang zwischen den beiden Variablen bestünde. Ein positives Residuum zeigt an, dass in der Zelle mehr Fälle vorliegen, als dies der Fall wäre, wenn die Zeilen- und Spaltenvariable unabhängig wären.
- **Angepasste Residuen.** Der Quotient aus dem Residuum einer Zelle (beobachteter Wert minus erwarteter Wert) und dessen geschätztem Standardfehler. Das resultierende standardisierte Residuum wird in Einheiten der Standardabweichung über oder unter dem Mittelwert angegeben.

Auswertungen für 2x2-Tabellen. In dieser Gruppe finden Sie Statistiken für Tabellen, in denen die Zeilen- und die Spaltenvariable jeweils zwei Kategorien aufweisen. Beide messen die Stärke des Zusammenhangs zwischen dem Vorhandensein eines Faktors und dem Auftreten eines Ereignisses.

- **Odds-Verhältnis.** Das Odds-Verhältnis kann als Schätzung des relativen Risikos verwendet werden, wenn der Faktor selten auftritt.
- **Relatives Risiko.** Das Verhältnis zwischen dem Risiko eines Ereignisses bei Vorliegen des Faktors zum Risiko des Ereignisses bei Fehlen des Faktors.
- **Risikodifferenz.** Die Differenz zwischen dem Risiko eines Ereignisses bei Vorliegen des Faktors zum Risiko des Ereignisses bei Fehlen des Faktors.

Test auf Unabhängigkeit von Zeilen und Spalten. Mit dieser Option werden Chi-Quadrat- und Likelihood-Quotienten-Tests für die Hypothese erstellt, dass eine Zeilen- und eine Spaltenvariable unabhängig sind. Für jedes Variablenpaar werden separate Tests durchgeführt.

Komplexe Stichproben: Fehlende Werte

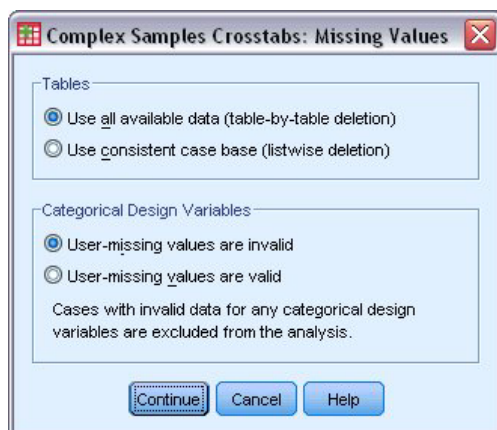


Abbildung 6. Dialogfeld "Fehlende Werte"

Tabellen. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Fehlende Werte werden jeweils für einzelne Tabellen bestimmt. Daher können die zur Berechnung von Statistiken verwendeten Fälle zwischen den einzelnen Häufigkeits- bzw. Kreuztabellen variieren.
- **Einheitliche Fallbasis verwenden.** Fehlende Werte werden über alle Variablen bestimmt. Daher sind die zur Berechnung der Statistiken verwendeten Fälle über die verschiedenen Tabellen konsistent.

Kategoriale Designvariablen. Diese Gruppe bestimmt, ob benutzerdefiniert fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

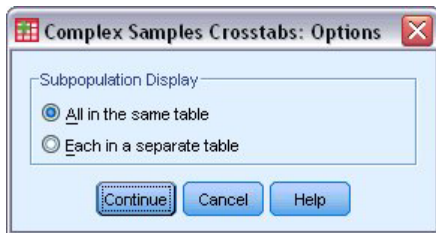


Abbildung 7. Dialogfeld "Optionen"

Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel 8. Verhältnisse für komplexe Stichproben

Die Prozedur "Verhältnisse für komplexe Stichproben" zeigt univariate Auswertungsstatistiken für Verhältnisse von Variablen an. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur "Verhältnisse für komplexe Stichproben" können Sie deskriptive Statistiken für das Verhältnis des aktuellen Eigenschaftswerts zum letzten bewerteten Wert erstellen. Diese Statistiken beruhen auf den Ergebnissen einer Studie, die anhand eines komplexen Designs und mit einem geeigneten Analyseplan für die Daten in einem US-Bundesstaat durchgeführt wurde.

Statistiken. Mit diesem Verfahren erhalten Sie Verhältnisschätzungen, *T*-Tests, Standardfehler, Konfidenzintervalle, Variationskoeffizienten, die ungewichteten Anzahlen, den Umfang der Grundgesamtheiten, die Effekte des Stichprobenplans und die Quadratwurzeln aus den Effekten des Stichprobenplans.

Erläuterungen der Daten für die Verhältnisse für komplexe Stichproben

Daten. Zähler und Nenner sollten metrische Variablen mit positivem Wert sein. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Ermitteln von Verhältnissen für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Verhältnisse...
2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
3. Klicken Sie auf **Weiter**.
4. Wählen Sie im Dialogfeld "Verhältnisse für komplexe Stichproben" mindestens eine Zählervariable und eine Nennervariable aus.

Optional können Sie Variablen zur Definition der Untergruppen angeben, für die Statistiken erstellt werden.

Verhältnisse für komplexe Stichproben: Statistiken

Statistiken. Mit dieser Gruppe erhalten Sie Statistiken zur Verhältnisschätzung.

- **Standardfehler.** Der Standardfehler der Schätzung.
- **Konfidenzintervall.** Ein Konfidenzintervall für die Schätzung unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers der Schätzung zur Schätzung.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung der Schätzung verwendet wurden.
- **Umfang der Grundgesamtheit.** Die geschätzte Anzahl an Einheiten in der Grundgesamtheit.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.

- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.

T-Test. Sie können *T*-Tests der Schätzungen unter Verwendung eines angegebenen Werts abrufen.

Verhältnisse für komplexe Stichproben: Fehlende Werte

Verhältnisse. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Fehlende Werte werden jeweils für einzelne Verhältnisse bestimmt. Daher können die zur Berechnung von Statistiken verwendeten Fälle zwischen den einzelnen Zähler/Nenner-Paaren variieren.
- **Einheitliche Fallbasis gewährleisten.** Fehlende Werte werden über alle Variablen bestimmt. Daher sind die zur Berechnung der Statistiken verwendeten Fälle konsistent.

Kategoriale Designvariablen. Diese Gruppe bestimmt, ob benutzerdefiniert fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

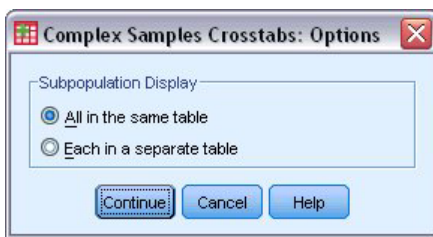


Abbildung 8. Dialogfeld "Optionen"

Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel 9. Allgemeines lineares Modell für komplexe Stichproben

Die Prozedur "Allgemeines lineares Modell für komplexe Stichproben" besteht aus einer linearen Regressionsanalyse sowie aus einer Analyse der Varianz und Kovarianz für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiel. Eine Lebensmittelkette hat eine Kundenumfrage über die Kaufgewohnheiten durchgeführt, die nach einem komplexen Design ausgeführt wurde. Auf der Grundlage der Umfrageergebnisse und der Zahlen über die Ausgaben der einzelnen Kunden im vergangenen Monat möchte das Unternehmen ermitteln, ob die Einkaufshäufigkeit in einem Zusammenhang mit den monatlichen Ausgaben steht, und zwar getrennt nach Geschlecht. Bei dieser Untersuchung soll das Stichprobendesign berücksichtigt werden.

Statistiken. Mit dieser Prozedur erhalten Sie Schätzungen, Standardfehler, Konfidenzintervalle, *t*-Tests, Effekte des Stichprobenplans und Quadratwurzeln aus den Effekten des Stichprobenplans, außerdem die Korrelationen und Kovarianzen bei den Parameterschätzungen. Auch Maße für die Anpassungsgüte des Modells und deskriptive Statistiken für die abhängigen und unabhängigen Variablen stehen zur Verfügung. Und nicht zuletzt können Sie geschätzte Randmittel für die Modellfaktorebenen und die Interaktionen zwischen Faktoren anfordern.

Erläuterungen der Daten beim allgemeinen linearen Modell für komplexe Stichproben

Daten. Die abhängige Variable ist quantitativ. Faktoren sind kategorial. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Erzeugen eines allgemeinen linearen Modells für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Allgemeines lineares Modell...
2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
3. Klicken Sie auf **Weiter**.
4. Wählen Sie im Dialogfeld "Allgemeines lineares Modell für komplexe Stichproben" eine abhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie die entsprechenden Variablen für Faktoren und Kovariaten gemäß den vorliegenden Daten aus.
- Geben Sie eine Variable an, um eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheitenvariable vorgenommen.

Allgemeines lineares Modell für komplexe Stichproben

Modelleffekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Interaktionseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

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

Haupteffekte. Legt einen Haupteffekterm für jede ausgewählte Variable an.

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

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

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

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

Verschachtelte Terme

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

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

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Interaktion müssen eindeutig sein. Dementsprechend ist die Angabe von A^*A unzulässig, wenn A ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von $A(A)$ unzulässig, wenn A ein Faktor ist.
- 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. Selbst wenn Sie den konstanten Term in das Modell aufnehmen, können Sie festlegen, dass die darauf bezogenen Statistiken unterdrückt werden sollen.

Allgemeines lineares Modell für komplexe Stichproben: Statistik

Modellparameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzung.** Zeigt eine Schätzung der Koeffizienten.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzungen.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzungen. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.

- **T-Test.** Zeigt je einen T -Test für die einzelnen Koeffizientenschätzungen. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzungen.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzungen.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.

Anpassungsgüte des Modells. Zeigt die Statistik für R^2 -Fehler und mittlere quadratische Fehler.

Mittelwerte der Grundgesamtheit für die abhängige Variable und die Kovariaten. Zeigt eine Zusammenfassung über die abhängige Variable, die Kovariaten und die Faktoren.

Informationen zum Stichprobendesign. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Hypothesentests für komplexe Stichproben

Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: F , F (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobendesign, mit denen die p -Werte für alle Teststatistiken berechnet werden. Dient das Stichprobendesign als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

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

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

Allgemeines lineares Modell für komplexe Stichproben: Geschätzte Mittelwerte

Im Dialogfeld "Geschätzte Mittelwerte" werden die vom Modell geschätzten Randmittel für die Ebenen der Faktoren und die Interaktionen zwischen Faktoren aufgeführt, die im untergeordneten Dialogfeld "Modell" angegeben wurden. Des Weiteren können Sie den Mittelwert für die gesamte Grundgesamtheit anzeigen lassen.

Term. Geschätzte Mittel werden für die ausgewählten Faktoren und Interaktionen zwischen Faktoren berechnet.

Kontrast. Der Kontrast bestimmt, wie die Hypothesentests zum Vergleich der geschätzten Mittel eingerichtet werden.

- *Einfach..* Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt.
- *Abweichung..* Vergleicht den Mittelwert jeder Faktorstufe (außer bei Referenzkategorien) mit dem Mittelwert aller Faktorstufen (Gesamtmittelwert). Die Stufen des Faktors können in beliebiger Ordnung vorliegen.
- *Differenz..* Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. Diese Kontraste werden auch als umgekehrte Helmert-Kontraste bezeichnet.
- *Helmert.* Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.
- *Wiederholt..* Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.
- *Polynomial.* Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

Referenzkategorie. Für einfache Kontraste und Abweichungskontraste wird eine Referenzkategorie oder eine Faktorebene benötigt, mit der die anderen Elemente verglichen werden können.

Allgemeines lineares Modell für komplexe Stichproben: Speichern

Variablen speichern. In dieser Gruppe speichern Sie die vom Modell vorhergesagten Werte und Residuen als neue Variablen in der Arbeitsdatei.

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

- **rowtype_.** Nimmt folgende Werte (und Wertbeschriftungen) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzungen), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobendesigns). 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 Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.

- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen 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 Parameterschätzungen sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Hinweis: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

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

Allgemeines lineares Modell für komplexe Stichproben: Optionen

Benutzerdefiniert fehlende Werte. Alle Designvariablen sowie die abhängige Variable und gegebenenfalls alle Kovariaten müssen gültige Daten enthalten. Fälle, bei denen ungültige Daten für diese Variablen vorliegen, werden aus der Analyse gelöscht. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Schicht-, Cluster-, Teilgesamtheits- und Faktorvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall. Dies ist die Konfidenzintervallebene für Koeffizientenschätzungen und geschätzte Randmittel. Geben Sie einen Wert größer oder gleich 50 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSGLM

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl **CUSTOM** können Sie benutzerdefinierte Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- 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 bei polynomialen Kontrasten eine Metrik angeben.
- Mit dem Unterbefehl **CRITERIA** können Sie einen Toleranzwert für die Prüfung auf Singularität festlegen.
- Mit dem Unterbefehl **SAVE** können Sie benutzerdefinierte Namen für gespeicherte Variablen angeben.
- Mit dem Unterbefehl **PRINT** können Sie eine Tabelle mit allgemeinen schätzbaren Funktionen anlegen.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 10. Logistische Regression für komplexe Stichproben

Die Prozedur "Logistische Regression für komplexe Stichproben" besteht aus einer logistischen Regressionsanalyse einer binären oder multinomialen abhängigen Variable für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiel. Ein Kreditsachbearbeiter verfügt über eine Reihe von Datensätzen zu Kunden, die ein Darlehen in verschiedenen Zweigstellen erhalten haben; diese Datensätze wurden nach einem komplexen Design zusammengestellt. Bei der Einbeziehung des Stichprobendesigns interessiert sich der Sachbearbeiter für die Wahrscheinlichkeit, mit der ein Kunde mit dem Darlehen in Verzug geraten könnte, und zwar im Zusammenhang mit dem Alter, der beruflichen Entwicklung und der Darlehenshöhe.

Statistiken. Mit dieser Prozedur erhalten Sie Schätzungen, potenzierte Schätzungen, Standardfehler, Konfidenzintervalle, *t*-Tests, Effekte des Stichprobenplans und Quadratwurzeln aus den Effekten des Stichprobenplans, außerdem die Korrelationen und Kovarianzen bei den Parameterschätzungen. Auch Pseudo- R^2 -Statistiken, Klassifizierungstabellen und deskriptive Statistiken für die abhängigen und unabhängigen Variablen stehen zur Verfügung.

Erläuterung der Daten bei der logistischen Regression für komplexe Stichproben

Daten. Die abhängige Variable ist kategorial. Faktoren sind kategorial. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Ermitteln der logistischen Regression für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Logistische Regression...
2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
3. Klicken Sie auf **Weiter**.
4. Wählen Sie im Dialogfeld "Logistische Regression für komplexe Stichproben" eine abhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie die entsprechenden Variablen für Faktoren und Kovariaten gemäß den vorliegenden Daten aus.
- Geben Sie eine Variable an, um eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheitenvariable vorgenommen.

Logistische Regression für komplexe Stichproben: Referenzkategorie

Bei der Prozedur "Logistische Regression für komplexe Stichproben" wird standardmäßig die höchstwertige Kategorie als Referenzkategorie verwendet. In diesem Dialogfeld können Sie den höchsten bzw. niedrigsten Wert oder auch eine benutzerdefinierte Kategorie als Referenzkategorie angeben.

Logistische Regression für komplexe Stichproben: Modell

Modelleffekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Interaktionseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

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

Haupteffekte. Legt einen Haupteffekterm für jede ausgewählte Variable an.

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

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

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

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

Verschachtelte Terme

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

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

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Interaktion müssen eindeutig sein. Dementsprechend ist die Angabe von $A*A$ unzulässig, wenn A ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von $A(A)$ unzulässig, wenn A ein Faktor ist.
- 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. Selbst wenn Sie den konstanten Term in das Modell aufnehmen, können Sie festlegen, dass die darauf bezogenen Statistiken unterdrückt werden sollen.

Logistische Regression für komplexe Stichproben: Statistik

Anpassungsgüte des Modells. Steuert die Anzeige der Statistik, in der die Gesamtleistung des Modells bewertet wird.

- **Pseudo-R-Quadrat.** Für die R^2 -Statistik aus der linearen Regression bieten die Modelle für die logistische Regression kein exaktes Gegenstück. Mit diesen Mehrfachmessungen werden stattdessen die Eigenschaften der R^2 -Statistik nachgebildet.

- **Klassifikationstabelle.** Zeigt die ausgewerteten Kreuzklassifikationen der beobachteten Kategorie nach der vom Modell vorhergesagten Kategorie für die abhängige Variable.

Parameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzung.** Zeigt eine Schätzung der Koeffizienten.
- **Potenzierte Schätzung.** Zeigt die Basis des natürlichen Logarithmus, potenziert mit den Schätzungen der Koeffizienten. Die Schätzung bietet zwar ergiebige Eigenschaften für statische Tests; die potenzierte Schätzung oder $\exp(B)$ ist jedoch einfacher zu interpretieren.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzungen.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzungen. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.
- **T-Test.** Zeigt je einen *T*-Test für die einzelnen Koeffizientenschätzungen. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzungen.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzungen.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.

Auswertungsstatistik für Modellvariablen. Zeigt eine Zusammenfassung über die abhängige Variable, die Kovariaten und die Faktoren.

Informationen zum Stichprobendesign. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Hypothesentests für komplexe Stichproben

Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: *F*, *F* (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

Stichprobenziehung; Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobendesign, mit denen die *p*-Werte für alle Teststatistiken berechnet werden. Dient das Stichprobendesign als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

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

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

- *Bonferroni sequenziell.* Hierbei handelt es sich um ein sequentielles 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.
- *Bonferroni.* Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.

Logistische Regression für komplexe Stichproben: Odds-Verhältnis

Im Dialogfeld "Odds-Verhältnisse" rufen Sie die vom Modell geschätzten Odds-Verhältnisse für bestimmte Faktoren und Kovariaten ab. Für jede Kategorie der abhängigen Variable (mit Ausnahme der Referenzkategorie) wird je ein separates Set von Odds-Verhältnissen berechnet.

Faktoren. Für jeden ausgewählten Faktor wird das Odds-Verhältnis in jeder Kategorie des Faktors zu den Quoten in der angegebenen Referenzkategorie angezeigt.

Kovariaten. Für jede ausgewählte Kovariate wird das Odds-Verhältnis für den Mittelwert der Kovariate zzgl. der angegebenen Änderungseinheiten zu den Quoten für den Mittelwert angezeigt.

Beim Berechnen der Odds-Verhältnisse für einen Faktor oder eine Kovariate werden alle anderen Faktoren auf die jeweils höchste Ebene fixiert, alle anderen Kovariaten dagegen auf den Mittelwert. Liegen Interaktionen eines Faktors oder einer Kovariate mit anderen Prädiktoren im Modell vor, sind die Odds-Verhältnisse nicht nur von den Änderungen bei der angegebenen Variable abhängig, sondern auch von den Werten der anderen Variablen, mit der interagiert wird. Wenn eine angegebene Kovariate im Modell Interaktionen mit sich selbst aufweist (z. B. *alter*alter*), sind die Odds-Verhältnisse sowohl von den Änderungen bei der Kovariate abhängig als auch vom Wert der Kovariate.

Logistische Regression für komplexe Stichproben: Speichern

Variablen speichern. In dieser Gruppe speichern Sie die vom Modell vorhergesagte Kategorie und die vorhergesagten Wahrscheinlichkeiten als neue Variablen im aktiven Dataset.

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

- **rowtype_.** Nimmt folgende Werte (und Wertbeschriftungen) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzungen), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobendesigns). 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 Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen 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 Parameterschätzungen sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Hinweis: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

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

Logistische Regression für komplexe Stichproben: Optionen

Schätzung. In dieser Gruppe steuern Sie verschiedene Kriterien für die Schätzung im Modell.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nicht negative Ganzzahl an.
- **Maximalzahl für Schritthalbierungen.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritthalbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationen auf der Grundlage der Änderung bei den Parameterschätzern begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzungen unter dem angegebenen (nicht negativen) Wert liegt.
- **Iterationen auf der Grundlage der Log-Likelihood-Änderung begrenzen.** 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 (nicht negativen) Wert liegt.
- **Prüfung auf vollständige Trennung der Datenpunkte.** Mit dieser Option lassen Sie Tests durch den Algorithmus durchführen, mit denen sichergestellt wird, dass die Parameterschätzungen eindeutige Werte aufweisen. Eine Trennung wird vorgenommen, sobald ein Modell erzeugt werden kann, in dem alle Fälle fehlerfrei klassifiziert werden.
- **Iterationsverlauf anzeigen.** Die Parameterschätzungen und die Statistik werden alle n Iterationen angezeigt, beginnend mit der 0. Iteration (den ursprünglichen Schätzungen). Wenn Sie den Iterationsverlauf drücken, wird die letzte Iteration stets unabhängig vom Wert für n ausgegeben.

Benutzerdefiniert fehlende Werte. Alle Designvariablen sowie die abhängige Variable und gegebenenfalls alle Kovariaten müssen gültige Daten enthalten. Fälle, bei denen ungültige Daten für diese Variablen vorliegen, werden aus der Analyse gelöscht. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Schicht-, Cluster-, Teilgesamtheits- und Faktorvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall. Dies ist die Konfidenzintervallebene für Koeffizientenschätzungen, potenzierte Koeffizientenschätzungen und Odds-Verhältnisse. Geben Sie einen Wert größer oder gleich 50 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSLOGISTIC

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl CUSTOM können Sie benutzerdefinierte Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl ODDS RATIOS können Sie Werte für andere Modellvariablen festlegen, wenn Sie die Odds-Verhältnisse für Faktoren und Kovariaten berechnen.
- Mit dem Unterbefehl CRITERIA können Sie einen Toleranzwert für die Prüfung auf Singularität festlegen.
- Mit dem Unterbefehl SAVE können Sie benutzerdefinierte Namen für gespeicherte Variablen angeben.
- Mit dem Unterbefehl PRINT können Sie eine Tabelle mit allgemeinen schätzbaren Funktionen anlegen.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 11. Ordinale Regression für komplexe Stichproben

Die Prozedur "Ordinale Regression für komplexe Stichproben" besteht aus einer Regressionsanalyse einer binären oder ordinalen abhängigen Variablen für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiel. Abgeordnete, die in Erwägung ziehen, einen Gesetzesentwurf einzubringen, sind daran interessiert zu ermitteln, ob dieser Gesetzesantrag öffentlich unterstützt wird und in welchem Bezug die Unterstützung für den Antrag zur demografischen Struktur der Wähler steht. Die Meinungsforscher verwenden für die Erstellung und Durchführung der entsprechenden Umfragen ein komplexes Stichprobendesign. Mit der ordinalen Regression für komplexe Stichproben können Sie ein Modell für die Stärke der Unterstützung für den Gesetzesentwurf auf der Grundlage der demografischen Struktur der Wähler anpassen.

Erläuterung der Daten bei der ordinalen Regression für komplexe Stichproben

Daten. Die abhängige Variable ist ordinal. Faktoren sind kategorial. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei den Variablen für die Teilgesamtheiten kann es sich um Zeichenfolgevariablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Ermitteln der ordinalen Regression für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren > Komplexe Stichproben > Ordinale Regression...

2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
3. Klicken Sie auf **Weiter**.
4. Wählen Sie im Dialogfeld "Ordinale Regression für komplexe Stichproben" eine abhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie die entsprechenden Variablen für Faktoren und Kovariaten gemäß den vorliegenden Daten aus.
- Geben Sie eine Variable an, um eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheitenvariable vorgenommen, die Varianzen jedoch werden dennoch ordnungsgemäß auf der Grundlage des gesamten Datensets geschätzt.
- Wählen Sie eine Verknüpfungsfunktion aus.

Verknüpfungsfunktion. Die Verknüpfungsfunktion ist eine Transformation der kumulativen Wahrscheinlichkeiten, die eine Schätzung des Modells ermöglicht. Die folgenden fünf Verknüpfungsfunktionen sind verfügbar:

- **Logit.** $f(x) = \log(x/(1-x))$. Wird üblicherweise für gleichmäßig verteilte Kategorien verwendet.
- **Log-Log komplementär.** $f(x) = \log(-\log(1-x))$. Wird üblicherweise verwendet, wenn höhere Kategorien wahrscheinlicher sind.
- **Log-Log negativ.** $f(x) = -\log(-\log(x))$. Wird üblicherweise verwendet, wenn niedrigere Kategorien wahrscheinlicher sind.

- **Probit.** $f(x) = \Phi^{-1}(x)$. Wird üblicherweise verwendet, wenn die latente Variable normalverteilt ist.
- **Cauchit (Inverse von Cauchy).** $f(x) = \tan(\pi(x-0.5))$. Wird üblicherweise verwendet, wenn die latente Variable viele Extremwerte aufweist.

Ordinale Regression für komplexe Stichproben: Antwortwahrscheinlichkeiten

Im Dialogfeld "Antwortwahrscheinlichkeiten" können Sie angeben, ob die kumulative Wahrscheinlichkeit einer Antwort (d. h. die Wahrscheinlichkeit der Zugehörigkeit bis zu einschließlich einer bestimmten Kategorie der abhängigen Variablen) mit steigenden bzw. sinkenden Werten der abhängigen Variablen steigt.

Ordinale Regression für komplexe Stichproben: Modell

Modelleffekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Interaktionseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

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

Haupteffekte. Legt einen Haupteffektterm für jede ausgewählte Variable an.

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

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

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

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

Verschachtelte Terme

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

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

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Interaktion müssen eindeutig sein. Dementsprechend ist die Angabe von $A*A$ unzulässig, wenn A ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von $A(A)$ unzulässig, wenn A ein Faktor ist.
- 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.

Ordinale Regression für komplexe Stichproben: Statistik

Anpassungsgüte des Modells. Steuert die Anzeige der Statistik, in der die Gesamtleistung des Modells bewertet wird.

- **Pseudo-R-Quadrat.** Für die R^2 -Statistik aus der linearen Regression bieten die Modelle für die ordinale Regression kein exaktes Gegenstück. Mit diesen Mehrfachmessungen werden stattdessen die Eigenschaften der R^2 -Statistik nachgebildet.
- **Klassifikationstabelle.** Zeigt die ausgewerteten Kreuzklassifikationen der beobachteten Kategorie nach der vom Modell vorhergesagten Kategorie für die abhängige Variable.

Parameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzung.** Zeigt eine Schätzung der Koeffizienten.
- **Potenzierte Schätzung.** Zeigt die Basis des natürlichen Logarithmus, potenziert mit den Schätzungen der Koeffizienten. Die Schätzung bietet zwar ergiebige Eigenschaften für statistische Tests; die potenzierte Schätzung oder $\exp(B)$ ist jedoch einfacher zu interpretieren.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzungen.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzungen. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.
- **T-Test.** Zeigt je einen T-Test für die einzelnen Koeffizientenschätzungen. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzungen.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzungen.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Designs (ausgedrückt in Einheiten, die denen des Standardfehlers vergleichbar sind). Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

Parallele Linien. In dieser Gruppe können Sie Statistiken anfordern, die einem Modell mit nicht parallelen Linien zugeordnet sind. Dabei wird eine separate Regressionslinie für jede Antwortkategorie (außer der letzten) angepasst.

- **Wald-Test.** Erstellt einen Test für die Nullhypothese, dass die Regressionsparameter für alle kumulativen Antworten gleich sind. Das Modell mit nicht parallelen Linien wird geschätzt und der Wald-Test auf gleiche Parameter wird angewendet.
- **Parameterschätzungen.** Zeigt Schätzwerte für die Koeffizienten und Standardfehler des Modells mit nicht parallelen Linien an.
- **Kovarianzen der Parameterschätzungen.** Zeigt eine Schätzung der Kovarianzmatrix für die Koeffizienten des Modells mit nicht parallelen Linien an.

Auswertungsstatistik für Modellvariablen. Zeigt eine Zusammenfassung über die abhängige Variable, die Kovariaten und die Faktoren.

Informationen zum Stichprobendesign. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Hypothesentests für komplexe Stichproben

Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: F , F (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobendesign, mit denen die p -Werte für alle Teststatistiken berechnet werden. Dient das Stichprobendesign als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

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

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

Ordinale Regression für komplexe Stichproben: Odds-Verhältnisse

Im Dialogfeld "Odds-Verhältnisse" rufen Sie die vom Modell geschätzten kumulativen Odds-Verhältnisse für bestimmte Faktoren und Kovariaten ab. Diese Funktion ist nur bei Modellen verfügbar, die die Verknüpfungsfunktion "Logit" verwenden. Für alle Kategorien der abhängigen Variablen mit Ausnahme der letzten wird ein einziges kumulatives Odds-Verhältnis berechnet; das proportionale Odds-Modell postuliert, dass alle gleich sind.

Faktoren. Für jeden ausgewählten Faktor wird das Verhältnis der kumulativen Quoten (Odds) in jeder Kategorie des Faktors zu den Quoten in der angegebenen Referenzkategorie angezeigt.

Kovariaten. Für jede ausgewählte Kovariate wird das Verhältnis der kumulativen Quoten (Odds) für den Mittelwert der Kovariaten zzgl. der angegebenen Änderungseinheiten zu den Quoten für den Mittelwert angezeigt.

Beim Berechnen der Odds-Verhältnisse für einen Faktor oder eine Kovariate werden alle anderen Faktoren auf die jeweils höchste Ebene fixiert, alle anderen Kovariaten dagegen auf den Mittelwert. Liegen Interaktionen eines Faktors oder einer Kovariate mit anderen Prädiktoren im Modell vor, sind die Odds-Verhältnisse nicht nur von den Änderungen bei der angegebenen Variable abhängig, sondern auch von den Werten der anderen Variablen, mit der interagiert wird. Wenn eine angegebene Kovariate im Modell Interaktionen mit sich selbst aufweist (z. B. *alter*alter*), sind die Odds-Verhältnisse sowohl von den Änderungen bei der Kovariate abhängig als auch vom Wert der Kovariate.

Ordinale Regression für komplexe Stichproben: Speichern

Variablen speichern. In dieser Gruppe speichern Sie die vom Modell vorhergesagte Kategorie, die Wahrscheinlichkeit der vorhergesagten Kategorie, die Wahrscheinlichkeit der beobachteten Kategorie und die vorhergesagten Wahrscheinlichkeiten als neue Variablen im aktiven Dataset.

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

- **rowtype_.** Nimmt folgende Werte (und Wertbeschriftungen) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzungen), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobendesigns). 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 Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen 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 Parameterschätzungen sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Hinweis: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

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

Ordinale Regression für komplexe Stichproben: Optionen

Schätzmethode. Sie können eine Methode für die Parameterschätzung auswählen. Sie haben die Wahl zwischen "Newton-Raphson", "Fisher-Scoring" und einer Hybridmethode, bei der zuerst Iterationen des Fisher-Scoring durchgeführt werden und dann zur Methode "Newton-Raphson" gewechselt wird. Wenn während der Phase "Fisher-Scoring" der Hybridmethode Konvergenz erreicht wird, bevor die maximale Anzahl an Fisher-Iterationen erreicht wurde, fährt der Algorithmus mit der Newton-Raphson-Methode fort.

Schätzung. In dieser Gruppe steuern Sie verschiedene Kriterien für die Schätzung im Modell.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nicht negative Ganzzahl an.
- **Maximalzahl für Schritthalbierungen.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritthalbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationen auf der Grundlage der Änderung bei den Parameterschätzern begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzungen unter dem angegebenen (nicht negativen) Wert liegt.
- **Iterationen auf der Grundlage der Log-Likelihood-Änderung begrenzen.** 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 (nicht negativen) Wert liegt.

- **Prüfung auf vollständige Trennung der Datenpunkte.** Mit dieser Option lassen Sie Tests durch den Algorithmus durchführen, mit denen sichergestellt wird, dass die Parameterschätzungen eindeutige Werte aufweisen. Eine Trennung wird vorgenommen, sobald ein Modell erzeugt werden kann, in dem alle Fälle fehlerfrei klassifiziert werden.
- **Iterationsverlauf anzeigen.** Die Parameterschätzungen und die Statistik werden alle n Iterationen angezeigt, beginnend mit der 0. Iteration (den ursprünglichen Schätzungen). Wenn Sie den Iterationsverlauf drucken, wird die letzte Iteration stets unabhängig vom Wert für n ausgegeben.

Benutzerdefiniert fehlende Werte. Metrische Designvariablen sowie die abhängige Variable und gegebenenfalls alle Kovariaten müssen gültige Daten enthalten. Fälle, bei denen ungültige Daten für diese Variablen vorliegen, werden aus der Analyse gelöscht. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Schicht-, Cluster-, Teilgesamtheits- und Faktorvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall. Dies ist die Konfidenzintervallebene für Koeffizientenschätzungen, potenzierte Koeffizientenschätzungen und Odds-Verhältnisse. Geben Sie einen Wert größer oder gleich 50 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSORDINAL

Die Befehlssyntax ermöglicht außerdem Folgendes:

- Mit dem Unterbefehl CUSTOM können Sie benutzerdefinierte Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl ODDS RATIOS können Sie andere Werte als die Mittelwerte für andere Modellvariablen festlegen, wenn Sie die kumulativen Odds-Verhältnisse für Faktoren und Kovariaten berechnen.
- Mit dem Unterbefehl ODDS RATIOS können Sie Werte ohne Beschriftung als benutzerdefinierte Referenzkategorien für Faktoren verwenden, wenn Odds-Verhältnisse angefordert werden.
- Mit dem Unterbefehl CRITERIA können Sie einen Toleranzwert für die Prüfung auf Singularität festlegen.
- Mit dem Unterbefehl PRINT können Sie eine Tabelle mit allgemeinen schätzbaren Funktionen anlegen.
- Mit dem Unterbefehl SAVE können Sie mehr als 25 Wahrscheinlichkeitsvariablen speichern.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Kapitel 12. Cox-Regression für komplexe Stichproben

Die Prozedur "Cox-Regression für komplexe Stichproben" bietet Funktionen zum Ausführen von Überlebensanalysen für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiele. Eine staatliche Strafverfolgungsbehörde befasst sich mit den Rückfallraten in ihrem Zuständigkeitsbereich. Einer der Messwerte für Rückfälligkeit ist die Zeit bis zur zweiten Festnahme von Straftätern. Die Behörde möchte anhand der Cox-Regression ein Modell für die Zeit bis zur erneuten Festnahme aufstellen, befürchtet aber, dass die Proportional-Hazards-Annahme für Alterskategorien nicht zutreffend sein könnte.

Medizinforscher untersuchen die Überlebenszeiten von Patienten nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls. Möglicherweise gibt es mehrere Fälle pro Patient, da die Anamnesen sich ändern, sobald das Auftreten von bedeutenden, nicht tödlichen Ereignissen vermerkt wird und die Zeitpunkte dieser Ereignisse festgehalten werden. Die Stichprobe ist außerdem insofern links gekürzt, als die beobachteten Überlebenszeiten durch die Dauer der Rehabilitation "überhöht" werden; während nämlich das Eintreten des Risikos zum Zeitpunkt des ischämischen Schlaganfalls beginnt, sind nur diejenigen Patienten Teil der Stichprobe, die bis zum Ende des Rehabilitationsprogramms überleben.

Erläuterung der Daten bei der Cox-Regression für komplexe Stichproben

Überlebenszeit. Bei der Prozedur wird die Cox-Regression auf Überlebenszeiten angewandt, also auf die Zeitspanne bis zum Eintreten eines Ereignisses. Es gibt abhängig von der Anfangszeit des Intervalls zwei Möglichkeiten, um die Überlebenszeit anzugeben:

- **Zeit=0.** Für gewöhnlich liegen Ihnen vollständige Informationen zum Beginn des Intervalls für jedes Subjekt vor und Sie haben eine Variable, die die Endzeiten enthält (oder Sie erstellen eine einzelne Variable mit Endzeiten aus Datums-/Zeitvariablen; siehe unten).
- **Variiert nach Subjekt.** Dies ist anzuwenden, wenn eine **Linksstutzung**, auch **Verzögerter Einschluss** genannt, vorliegt; falls Sie zum Beispiel die Überlebenszeiten für Patienten nach einem Rehabilitationsprogramm analysieren, das nach einem Schlaganfall durchgeführt wurde, könnten Sie den Zeitpunkt des Schlaganfalls mit dem Eintreten des Risikos gleichsetzen. Falls bei Ihrer Stichprobe jedoch nur Patienten berücksichtigt werden, die das Rehabilitationsprogramm überlebt haben, ist sie insofern links gekürzt, als die beobachteten Überlebenszeiten durch die Dauer der Rehabilitation "überhöht" werden. Sie können dem Rechnung tragen, indem Sie den Abschlusszeitpunkt der Rehabilitation als den Zeitpunkt des Einschlusses in die Studie angeben.

Datums- und Zeitvariablen. Datums- und Zeitvariablen können nicht zur direkten Bestimmung des Anfangs und des Endes des Intervalls verwendet werden; Datums- und Zeitvariablen sollten verwendet werden, um Variablen zu erstellen, die Überlebenszeiten enthalten. Falls keine Linksstutzung vorliegt, erstellen Sie einfach eine Variable, die Endzeiten auf Basis der Differenz zwischen dem Zeitpunkt des Eintritts in die Studie und dem Beobachtungsdatum enthält. Falls eine Linksstutzung vorliegt, erstellen Sie eine Variable, die Anfangszeiten auf Basis der Differenz zwischen dem Anfangsdatum der Studie und dem Einschlussdatum enthält, sowie eine Variable, die Endzeiten auf Basis der Differenz zwischen dem Anfangsdatum der Studie und dem Beobachtungsdatum enthält.

Ereignisstatus. Sie benötigen eine Variable, die erfasst, ob das Subjekt das relevante Ereignis innerhalb des Intervalls erfahren hat. Subjekte, bei denen das Ereignis nicht eingetreten ist, sind rechtszensiert.

Subjekt-ID. Sie können stückweise konstante, zeitabhängige Prädiktoren einfach integrieren, indem Sie die Beobachtungen zu einem einzelnen Subjekt auf mehrere Fälle aufteilen. Wenn Sie zum Beispiel Überlebenszeiten für Patienten nach einem Schlaganfall analysieren, können die Variablen, die für ihre Anam-

nese stehen, als Prädiktoren verwendet werden. Im Laufe der Zeit ereignen sich unter Umständen bedeutende medizinische Ereignisse, die ihre Anamnese ändern. Die folgende Tabelle zeigt die Strukturierung eines solchen Datensets: *Patienten-ID* ist die Subjekt-ID, *Endzeit* legt die beobachteten Intervalle fest, *Status* erfasst bedeutende medizinische Ereignisse und *Herzinfarktanamnese* und *Blutungsanamnese* sind stückweise konstante, zeitabhängige Prädiktoren.

Tabelle 1. Datenstruktur für die Integration stückweise konstanter, zeitabhängiger Prädiktoren.

Patienten-ID	Endzeit	Status	Herzinfarktanamnese	Blutungsanamnese
E	5	Herzinfarkt	Nein	Nein
E	7	Blutung	Ja	Nein
E	8	Verstorben	Ja	Ja
Z	24	Verstorben	Nein	Nein
3	8	Herzinfarkt	Nein	Nein
3	15	Verstorben	Ja	Nein

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Design, dessen Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Normalerweise wird bei Cox-Regressionsmodellen von proportionalen Hazards ausgegangen; das heißt, dass sich das Hazardverhältnis von einem Fall zum nächsten über die Zeit nicht ändern darf. Falls diese Annahme nicht zutreffend ist, müssen Sie dem Modell unter Umständen zeitabhängige Prädiktoren hinzufügen.

Kaplan-Meier-Analyse. Falls Sie keine Prädiktoren auswählen (oder keine ausgewählten Prädiktoren in das Modell eingeben) und für die Berechnung der Basisüberlebenskurve auf der Registerkarte "Optionen" die Produkt-Limit-Methode auswählen, wird eine Überlebensanalyse nach der Kaplan-Meier-Methode durchgeführt.

So erstellen Sie eine Cox-Regression für komplexe Stichproben

1. Wählen Sie die folgenden Befehle aus den Menüs aus:
Analysieren > Komplexe Stichproben > Cox-Regression...
2. Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
3. Klicken Sie auf **Weiter**.
4. Geben Sie die Überlebenszeit an, indem Sie den Zeitpunkt des Einschlusses in die Studie und den Zeitpunkt des Austritts aus der Studie auswählen.
5. Wählen Sie eine Ereignisstatusvariable aus.
6. Klicken Sie auf Ereignis definieren und legen Sie mindestens einen Ereigniswert fest.

Wahlweise können Sie eine Subjekt-ID auswählen.

Ereignis definieren

Geben Sie die Werte ein, die anzeigen, dass ein terminales Ereignis aufgetreten ist.

- **Individuelle(r) Wert(e).** Geben Sie eine oder mehrere Variablen an, indem Sie sie im Raster eingeben oder aus einer Liste mit definierten Wertbeschriftungen auswählen.
- **Wertebereich.** Geben Sie einen Wertebereich an, indem Sie die Minimal- und Maximalwerte eingeben oder Werte aus einer Liste mit definierten Wertbeschriftungen auswählen.

Prädiktoren

Auf der Registerkarte "Prädiktoren" können Sie die Faktoren und Kovariaten angeben, die zum Erstellen der Modelleffekte verwendet werden.

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

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

Zeitabhängige Prädiktoren. In bestimmten Situationen ist die Proportional-Hazards-Annahme nicht zutreffend. Dies bedeutet, dass sich Hazardverhältnisse im Laufe der Zeit ändern: Die Werte einer (oder mehrerer) Prädiktoren sind zu verschiedenen Zeitpunkten unterschiedlich. In solchen Fällen müssen Sie zeitabhängige Prädiktoren angeben. Weitere Informationen hierzu finden Sie unter „Zeitabhängige Prädiktoren definieren“. Zeitabhängige Prädiktoren können als Faktoren oder als Kovariaten angegeben werden.

Zeitabhängige Prädiktoren definieren

Im Dialogfeld "Zeitabhängigen Prädiktor definieren" können Sie einen Prädiktor definieren, der von der systemeigenen Zeitvariablen $T_$ abhängig ist. Sie können diese Variable verwenden, um zeitabhängige Kovariaten auf zwei Arten zu definieren.

- Wenn Sie ein erweitertes Cox-Regressionsmodell schätzen wollen, mit dem nicht proportionale Hazards möglich sind, definieren Sie den zeitabhängigen Prädiktor als eine Funktion der Zeitvariablen $T_$ und der fraglichen Kovariate. Ein geläufiges Beispiel wäre das einfache Produkt aus Zeitvariable und Prädiktor, aber es können auch komplexere Funktionen festgelegt werden.
- Einige Variablen können zu unterschiedlichen Zeiten unterschiedliche Werte aufweisen, weisen aber kein systematisches Verhältnis zur Zeit auf. In solchen Fälle müssen Sie einen **segmentierten zeitabhängigen Prädiktor** 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 den zeitabhängigen Prädiktor aus einem Set 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 den zeitabhängigen Prädiktor 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.

Hinweis: Falls der segmentierte, zeitabhängige Prädiktor innerhalb von Segmenten so wie im oben aufgeführten Blutdruckbeispiel konstant ist, ist es möglicherweise einfacher, einen stückweise konstanten, zeitabhängigen Prädiktor anzugeben, indem Sie Subjekte auf mehrere Fälle aufteilen. In der Beschreibung von Subjekt-IDs in Kapitel 12, „Cox-Regression für komplexe Stichproben“, auf Seite 49 finden Sie weitere Informationen.

Sie können die Steuerelemente im Dialogfeld "Zeitabhängigen Prädiktor definieren" verwenden, um den Ausdruck für die zeitabhängige Kovariate zu bilden, oder Sie können diesen direkt in das Eingabefeld "Numerischer Ausdruck" eingeben. Beachten Sie, dass Zeichenfolgekonstanten in Anführungszeichen oder Apostrophe gesetzt und numerische Konstanten in amerikanischem Format mit einem Punkt als Dezimaltrennzeichen eingegeben werden müssen. Die daraus resultierende Variable erhält den von Ihnen angegebenen Namen und sollte als Faktor oder als Kovariate auf der Registerkarte "Prädiktoren" enthalten sein.

Untergruppen

Basisschichten Für jeden Wert dieser Variablen wird eine separate Basishazard- und Basisüberlebensfunktion berechnet, während über alle Schichten hinweg ein einziges Set von Modellkoeffizienten geschätzt wird.

Teilgesamtheitenvariable. Geben Sie eine Variable an, um eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheitenvariable vorgenommen.

Modell

Modelleffekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Interaktionseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

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

Haupteffekte. Legt einen Haupteffektterm für jede ausgewählte Variable an.

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

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

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

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

Verschachtelte Terme

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

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

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Interaktion müssen eindeutig sein. Dementsprechend ist die Angabe von $A*A$ unzulässig, wenn A ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von $A(A)$ unzulässig, wenn A ein Faktor ist.
- 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.

Statistik

Informationen zum Stichprobendesign. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Zusammenfassung Ereignis und Zensur. Zeigt zusammenfassende Angaben zur Anzahl und den Prozentsatz von zensierten Fällen an.

Risikoset an Ereigniszeitpunkten. Zeigt die Anzahl von Ereignissen und die Anzahl mit Risiko für jeden Ereigniszeitpunkt in jeder Basisschicht an.

Parameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzung.** Zeigt eine Schätzung der Koeffizienten.
- **Potenzierte Schätzung.** Zeigt die Basis des natürlichen Logarithmus, potenziert mit den Schätzungen der Koeffizienten. Die Schätzung bietet zwar ergiebige Eigenschaften für statische Tests; die potenzierte Schätzung oder $\exp(B)$ ist jedoch einfacher zu interpretieren.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzungen.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzungen. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.
- **T-Test.** Zeigt je einen T -Test für die einzelnen Koeffizientenschätzungen. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzungen.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzungen.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Designeffekt.** Das Verhältnis der Varianz der Schätzung zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.
- **Quadratwurzel aus dem Effekt des Stichprobendesigns.** Dies ist ein Maß für den Effekt eines komplexen Stichprobenplans, wobei eine größere Abweichung von 1 auf größere Effekte hinweist.

Modellannahmen. Diese Gruppe ermöglicht Ihnen die Durchführung eines Tests der Proportional-Hazards-Annahme. Der Test vergleicht das angepasste Modell mit einem alternativen Modell, das zeitabhängige Prädiktoren x^*_{TF} für jeden Prädiktor x enthält, wobei $_{TF}$ die angegebene Zeitfunktion ist.

- **Zeitfunktion.** Gibt die Form von $_{TF}$ für das Alternativmodell an. Für die **Identitäts**-Funktion gilt: $_{TF}=T_.$ Für die **Log**-Funktion gilt: $_{TF}=\log(T_)$. Für **Kaplan-Meier** gilt: $_{TF}=1-S_{KM}(T_)$, wobei $S_{KM}(\cdot)$ die Kaplan-Meier-Schätzung der Überlebensfunktion ist. Bei **Rangist** $_{TF}$ die Rangordnung von $T_$ der beobachteten Endzeiten.
- **Parameterschätzungen für das Alternativmodell.** Zeigt die Schätzung, den Standardfehler sowie das Konfidenzintervall für jeden Parameter im Alternativmodell an.
- **Kovarianzmatrix für das Alternativmodell.** Zeigt die Matrix für geschätzte Kovarianzen zwischen Parametern im Alternativmodell an.

Basisüberlebens- und kumulative Basishazardfunktionen. Zeigt die Basisüberlebensfunktion und die kumulative Basishazardfunktion sowie die dazugehörigen Standardfehler an.

Hinweis: Falls zeitabhängige Prädiktoren, die auf der Registerkarte "Prädiktoren" definiert sind, im Modell enthalten sind, ist diese Option nicht verfügbar.

Diagramme

Die Registerkarte "Diagramme" ermöglicht es Ihnen, Diagramme der Hazardfunktion, der Überlebensfunktion, der Log-minus-Log der Überlebensfunktion und der Eins-minus-Überlebensfunktion anzufordern. Sie können außerdem auswählen, Konfidenzintervalle der angegebenen Funktionen als Diagramm darzustellen; das Konfidenzniveau wird auf der Registerkarte "Optionen" eingestellt.

Prädiktormuster. Sie können ein Muster von Prädiktorwerten angeben, das für die angeforderten Diagramme und für die exportierte Überlebens-Datei verwendet wird, die über die Registerkarte "Export" er-

stellt wird. Beachten Sie, dass diese Optionen nicht verfügbar sind, falls zeitabhängige Prädiktoren, die auf der Registerkarte "Prädiktoren" definiert sind, im Modell enthalten sind.

- **Faktoren im Diagramm darstellen in.** In der Standardeinstellung wird jeder Faktor auf der höchsten Ebene ausgewertet. Falls erwünscht, können Sie eine andere Ebene eingeben oder auswählen. Alternativ können Sie festlegen, dass für jede Ebene eines einzelnen Faktors separate Linien erzeugt werden, indem Sie auf das Kontrollkästchen für diesen Faktor klicken.
- **Kovariaten im Diagramm darstellen in.** Jede Kovariate wird an ihrem Mittelwert ausgewertet. Falls erwünscht, können Sie einen anderen Wert eingeben oder auswählen.

Hypothesentests

Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: F , F (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobendesign, mit denen die p -Werte für alle Teststatistiken berechnet werden. Dient das Stichprobendesign als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

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

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

Speichern

Variablen speichern. In dieser Gruppe speichern Sie modellbezogene Variablen im aktiven Dataset für die weitere Verwendung bei Diagnosen und Ergebnisberichten. Beachten Sie, dass keine dieser Optionen verfügbar ist, falls zeitabhängige Prädiktoren im Modell enthalten sind.

- **Überlebensfunktion.** Speichert die Überlebenswahrscheinlichkeit (den Wert der Überlebensfunktion) zum Zeitpunkt der Beobachtung sowie Prädiktorwerte für jeden Fall.
- **Untergrenze des Konfidenzintervalls für die Überlebensfunktion.** Speichert die Untergrenze des Konfidenzintervalls für die Überlebensfunktion zum Zeitpunkt der Beobachtung sowie Prädiktorwerte für jeden Fall.
- **Obergrenze des Konfidenzintervalls für die Überlebensfunktion.** Speichert die Obergrenze des Konfidenzintervalls für die Überlebensfunktion zum Zeitpunkt der Beobachtung sowie Prädiktorwerte für jeden Fall.
- **Kumulative Hazardfunktion.** Speichert die kumulative Hazardfunktion, oder $-\ln(\text{survival})$, zum Zeitpunkt der Beobachtung sowie Prädiktorwerte für jeden Fall.

- **Untergrenze des Konfidenzintervalls für die kumulative Hazardfunktion.** Speichert die Untergrenze des Konfidenzintervalls für die kumulative Hazardfunktion zum Zeitpunkt der Beobachtung sowie Prädiktorwerte für jeden Fall.
- **Obergrenze des Konfidenzintervalls für die kumulative Hazardfunktion.** Speichert die Obergrenze des Konfidenzintervalls für die kumulative Hazardfunktion zum Zeitpunkt der Beobachtung sowie Prädiktorwerte für jeden Fall.
- **Vorhergesagter Wert für den linearen Prädiktor.** Speichert die lineare Kombination von per Referenzwert korrigierten Prädiktoren mal Regressionskoeffizienten. Der lineare Prädiktor ist das Verhältnis der Hazardfunktion zum Basishazard. Im Proportional-Hazards-Modell ist dieser Wert über die Zeit konstant.
- **Schoenfeld-Residuum.** Für jeden nicht zensierten Fall und jeden nicht redundanten Parameter im Modell ist das Schoenfeld-Residuum die Differenz zwischen dem beobachteten Wert für den Prädiktor, der dem Modellparameter zugeordnet ist, und dem erwarteten Wert für Fälle, die sich zum beobachteten Zeitpunkt im Risikoset befinden. Schoenfeld-Residuen können bei der Bewertung der Proportional-Hazards-Annahme verwendet werden; für einen Prädiktor x zum Beispiel zeigen Diagramme der Schoenfeld-Residuen für den zeitabhängigen Prädiktor $x \cdot \ln(T_0)$ über Zeit eine horizontale Linie bei 0, falls die proportionale Hazardannahme zutrifft. Für jeden nicht redundanten Parameter im Modell wird eine separate Variable gespeichert. Schoenfeld-Residuen werden nur für unzensierte Fälle berechnet.
- **Martingale-Residuum.** Für jeden Fall ist das Martingale-Residuum die Differenz zwischen der beobachteten Zensur (0 falls zensiert, 1 falls nicht) und der Erwartung eines Ereignisses während des Beobachtungszeitraums.
- **Abweichungsresiduum.** Abweichungsresiduen sind Martingale-Residuen, die "korrigiert" wurden, um bei 0 symmetrischer zu erscheinen. Diagramme von Abweichungsresiduen in Abhängigkeit von Prädiktoren sollten keine Muster aufweisen.
- **Cox-Snell-Residuum.** Für jeden Fall ist das Cox-Snell-Residuum die Erwartung eines Ereignisses während des Beobachtungszeitraums oder die beobachtete Zensur minus das Martingale-Residuum.
- **Scoreresiduum.** Für jeden Fall oder jeden nicht redundanten Parameter im Modell ist das Scoreresiduum der Anteil des Falls an der ersten Ableitung der Pseudo-Likelihood. Für jeden nicht redundanten Parameter im Modell wird eine separate Variable gespeichert.
- **DFBeta-Residuum.** Für jeden Fall oder jeden nicht redundanten Parameter im Modell ist das DFBeta-Residuum der Näherungswert der Änderung im Wert der Parameterschätzung, der eintritt, wenn der Fall aus dem Modell entfernt wird. Fälle mit relativ großen DFBeta-Residuen können einen übermäßigen Einfluss auf die Analyse ausüben. Für jeden nicht redundanten Parameter im Modell wird eine separate Variable gespeichert.
- **Aggregierte Residuen.** Wenn mehrere Fälle für ein einzelnes Subjekt stehen, ist das aggregierte Residuum für ein Subjekt einfach die Summe der jeweiligen Fallresiduen aus allen Fällen, die zum selben Subjekt gehören. Beim Schoenfeld-Residuum unterscheidet sich die aggregierte Version nicht von der nicht aggregierten Version, da das Schoenfeld-Residuum nur für unzensierte Fälle definiert ist. Diese Residuen sind nur dann verfügbar, wenn eine Subjekt-ID auf der Registerkarte "Zeit und Ereignis" angegeben ist.

Namen der gespeicherten Variablen. Durch eine automatische Generierung von Namen wird sichergestellt, dass Ihre Arbeit nicht verloren geht. Mit benutzerdefinierten Namen können Sie Ergebnisse aus früheren Durchgängen verwerfen/ersetzen, ohne zuerst die gespeicherten Variablen im Dateneditor löschen zu müssen.

Export

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

- **rowtype_.** Nimmt folgende Werte (und Wertbeschriftungen) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzungen), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobendesigns). 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 Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertbeschriftungen, die den in der Tabelle der Parameterschätzungen angezeigten Parameterzeichenfolgen 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 Parameterschätzungen sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Hinweis: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

Überlebensfunktion als IBM SPSS Statistics Daten exportieren. Schreibt ein Dataset im IBM SPSS Statistics-Format, das Folgendes enthält: die Überlebensfunktion, den Standardfehler der Überlebensfunktion, Ober- und Untergrenzen des Konfidenzintervalls der Überlebensfunktion und die kumulative Hazardfunktion für jeden Versagens- oder Ereigniszeitpunkt, der an der Basis und nach den auf der Registerkarte "Diagramme" angegebenen Prädiktormustern ausgewertet wird. Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **Geschichtete Basisvariable.** Für jeden Wert der geschichteten Variablen werden separate Überlebenstabellen erzeugt.
- **Überlebenszeitvariable.** Der Ereigniszeitpunkt; ein separater Fall wird für jeden einzigartigen Ereigniszeitpunkt angelegt.
- **Sur_0, LCL_Sur_0, UCL_Sur_0.** Die Basisüberlebensfunktion und die Ober- und Untergrenzen ihres Konfidenzintervalls.
- **Sur_R, LCL_Sur_R, UCL_Sur_R.** Die am "Referenzmuster" ausgewertete Überlebensfunktion (siehe die Tabelle der Musterwerte in der Ausgabe) und die Ober- und Untergrenzen ihres Konfidenzintervalls.
- **Sur_##, LCL_Sur_##, UCL_Sur_##, ...** Überlebensfunktion, die an jedem auf der Registerkarte "Diagramme" angegebenen Prädiktormuster ausgewertet wird, und die Ober- und Untergrenzen ihres Konfidenzintervalls. Verwenden Sie die Tabelle der Musterwerte in der Ausgabe, um die Muster mit der Nummer ## zu vergleichen.
- **Haz_0, LCL_Haz_0, UCL_Haz_0.** Die kumulative Basishazardfunktion und die Ober- und Untergrenzen ihres Konfidenzintervalls.
- **Haz_R, LCL_Haz_R, UCL_Haz_R.** Die am "Referenzmuster2" ausgewertete kumulative Hazardfunktion (siehe die Tabelle der Musterwerte in der Ausgabe) und die Ober- und Untergrenzen ihres Konfidenzintervalls.
- **Haz_##, LCL_Haz_##, UCL_Haz_##, ...** Kumulative Hazardfunktion, die an jedem auf der Registerkarte "Diagramme" angegebenen Prädiktormuster ausgewertet wird, und die Ober- und Untergrenzen ihres Konfidenzintervalls. Verwenden Sie die Tabelle der Musterwerte in der Ausgabe, um die Muster mit der Nummer ## zu vergleichen.

Modell als XML exportieren. Speichert alle Informationen, die zur Vorhersage der Überlebensfunktion nötig sind, sowie Parameterschätzungen und die Basisüberlebensfunktion im XML-Format (PMML). Anhand dieser Modelldatei können Sie die Modellinformationen zu Scoring-Zwecken auf andere Datendateien anwenden.

Optionen

Schätzung. Diese Steuerelemente geben die Kriterien für die Schätzung von Regressionskoeffizienten an.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nicht negative Ganzzahl an.
- **Maximalzahl für Schritthalbierungen.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritthalbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationen auf der Grundlage der Änderung bei den Parameterschätzern begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzungen unter dem angegebenen (positiven) Wert liegt.
- **Iterationen auf der Grundlage der Log-Likelihood-Änderung begrenzen.** 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.
- **Iterationsverlauf anzeigen.** Zeigt den Iterationsverlauf für Parameterschätzungen und Pseudo-Likelihood an und druckt die letzte Evaluierung der Änderung der Parameterschätzungen und Pseudo-Log-Likelihood. Die Tabelle mit dem Iterationsverlauf druckt jede n te Iteration, beginnend mit der n ten Iteration (Anfangsschätzungen), wobei n das Inkrement ist. Wenn der Iterationsverlauf angefordert wird, wird die letzte Iteration stets angezeigt, unabhängig von n .
- **Entscheidungsmethode für die Parameterschätzung.** Wenn Bindungen zwischen beobachteten Versagenszeitpunkten bestehen, wird eine dieser Methoden angewandt, um die Bindungen zu lösen. Die Efron-Methode erfordert eine höhere Rechenleistung.

Überlebensfunktionen. Diese Steuerelemente geben die Kriterien für Berechnungen im Zusammenhang mit der Überlebensfunktion an.

- **Methode zur Schätzung von Basisüberlebensfunktionen.** Die **Breslow**-Methode (oder die **Nelson-Aalan**-Methode oder die empirische Methode) schätzt den Grundwert für die kumulative Hazardrate durch eine nicht abnehmende Schrittfunktion mit Schritten bei den beobachteten Versagenszeitpunkten und berechnet daraufhin die Basisüberlebensrate anhand der Beziehung $\text{Überleben} = \exp(-\text{kumulative Hazardrate})$. Die **Efron**-Methode ist rechenintensiver und beschränkt sich auf die **Breslow**-Methode, falls keine Bindungen bestehen. Die **Produkt-Limit**-Methode schätzt die Basisüberlebensrate anhand einer nicht steigenden rechtsstetigen Funktion; wenn das Modell keine Prädiktoren enthält, beschränkt sich diese Methode auf die **Kaplan-Meier**-Schätzung.
- **Konfidenzintervalle von Überlebensfunktionen.** Das Konfidenzintervall kann auf drei verschiedene Arten berechnet werden: in ursprünglichen Einheiten, über eine logarithmische Transformation oder über eine **Log-minus-Log**-Transformation. Nur die **Log-minus-Log**-Transformation garantiert, dass die Grenzen des Konfidenzintervalls zwischen 0 und 1 liegen, dennoch bringt die logarithmische Transformation im Allgemeinen die "besten" Ergebnisse.

Benutzerdefiniert fehlende Werte. Alle Variablen 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 kategorischen Modellen (wie Faktor-, Ereignis-, Schicht- und Teilgesamtheitsvariablen) und Stichprobendesignvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall(%). Dies ist die Konfidenzintervallebene für Koeffizientenschätzungen, potenzierte Koeffizientenschätzungen, Überlebensfunktionsschätzungen und Schätzungen der kumulativen Hazardfunktion. Geben Sie einen Wert größer oder gleich 0 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSCOXREG

Mit der Befehlssprache verfügen Sie außerdem über folgende Möglichkeiten:

- Benutzerdefinierte Hypothesentests durchführen (unter Verwendung des Unterbefehls CUSTOM und /PRINT LMATRIX).
- Toleranzspezifikation (unter Verwendung von /CRITERIA SINGULAR).
- Tabelle mit allgemein schätzbaren Funktionen (unter Verwendung von /PRINT GEF).
- Muster aus mehreren Prädiktoren (unter Verwendung von mehreren PATTERN-Unterbefehlen).
- Maximale Anzahl von gespeicherten Variablen, wenn ein Stammname angegeben wurde (unter Verwendung des Unterbefehls SAVE). Das Dialogfeld berücksichtigt den CSCOXREG-Standard von 25 Variablen.

Vollständige Informationen zur Syntax finden Sie in der Befehlssyntaxreferenz.

Bemerkungen

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden.

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

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

IBM Director of Licensing
IBM Europe, Middle East & Africa
Tour Descartes
2, avenue Gambetta
92066 Paris La Defense
France

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

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

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

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
USA

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

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

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können davon abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

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

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

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

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

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

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

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

© Copyright IBM Corp. _Jahr/Jahre angeben_. Alle Rechte vorbehalten.

Marken

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corporation in den USA und/oder anderen Ländern. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „Copyright and trademark information“ unter www.ibm.com/legal/copytrade.shtml.

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

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

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

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

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

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

Index

A

- Abweichungskontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34
- Abweichungsresiduen
 - in Cox-Regression für komplexe Stichproben 54
- Aggregierte Residuen
 - in Cox-Regression für komplexe Stichproben 54
- Allgemeines lineares Modell für komplexe Stichproben 31
 - geschätzte Mittel 34
 - Modell 32
 - Optionen 35
 - Statistik 32
 - Variablen speichern 34
 - zusätzliche Funktionen beim Befehl 35
- Analyseplan 13
- Antwortwahrscheinlichkeiten
 - bei der ordinalen Regression für komplexe Stichproben 44

B

- Basisschichten
 - in Cox-Regression für komplexe Stichproben 52
- Bonferroni
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Breslow-Schätzmethode
 - in Cox-Regression für komplexe Stichproben 57
- Brewers Stichprobenmethode
 - beim Stichprobenassistenten 6

C

- Chi-Quadrat
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Cluster
 - beim Stichprobenassistenten 5
 - im Analysevorbereitungsassistenten 13
- Cox-Regression für komplexe Stichproben
 - Datum- und Uhrzeitvariablen 49
 - Ereignis definieren 50
 - Grafik 53
 - Hypothesentests 54
 - Kaplan-Meier-Analyse 49
 - Modell 52
 - Modellexport 55
 - Optionen 57
 - Prädiktoren 51
 - Statistik 52

- Cox-Regression für komplexe Stichproben (*Forts.*)
 - Untergruppen 52
 - Variablen speichern 54
 - zeitabhängiger Prädiktor 51
- Cox-Snell-Residuen
 - in Cox-Regression für komplexe Stichproben 54

D

- Designeffekt
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Cox-Regression für komplexe Stichproben 52
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29
- Deskriptive Statistiken für komplexe Stichproben 23
 - fehlende Werte 24
 - Statistik 23
- Differenzkontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34

E

- Efron-Schätzmethode
 - in Cox-Regression für komplexe Stichproben 57
- Einfache Kontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34
- Einfache Zufallsstichprobenziehung
 - beim Stichprobenassistenten 6
- Eingabestichprobengewichtung
 - beim Stichprobenassistenten 5
- Einschlusswahrscheinlichkeiten
 - beim Stichprobenassistenten 8
- Erwartete Werte
 - in Kreuztabellen für komplexe Stichproben 25

F

- F (angepasst), Statistik
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46

- F-Statistik
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Fehlende Werte
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47
 - im allgemeinen linearen Modell für komplexe Stichproben 35
 - in "Deskriptive Statistiken für komplexe Stichproben" 24
 - in komplexen Stichproben 20, 26
 - in Verhältnisse für komplexe Stichproben 30
- Fisher-Scoring
 - bei der ordinalen Regression für komplexe Stichproben 47
- Freiheitsgrade
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46

G

- Geringste signifikante Differenz
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Geschätzte Randmittel
 - im allgemeinen linearen Modell für komplexe Stichproben 34

H

- Häufigkeiten für komplexe Stichproben 19
 - Statistik 19
- Helmert-Kontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34

I

- Informationen zum Stichprobendesign
 - in Cox-Regression für komplexe Stichproben 52
- Iteration
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47
- Iterationsverlauf
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47

K

- Klassifikationstabellen
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
- Komplexe Stichproben
 - Analyseplan 13
 - fehlende Werte 20, 26
 - Hypothesentests 33, 39, 46
 - Optionen 21, 24, 27, 30
 - Stichprobenplan 5
- Konfidenzintervalle
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32, 35
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29
- Konfidenzniveau
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47
- Kontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34
- Korrelationen der Parameterschätzungen
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
- Korrigierte Residuen
 - in Kreuztabellen für komplexe Stichproben 25
- Korrigiertes Chi-Quadrat
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Kovarianzen der Parameterschätzungen
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
- Kreuztabellen für komplexe Stichproben 25
 - Statistik 25
- Kumulative Wahrscheinlichkeiten
 - bei der ordinalen Regression für komplexe Stichproben 47
- Kumulative Werte
 - in "Häufigkeiten für komplexe Stichproben" 19

L

- Likelihood-Konvergenz
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47
- Logistische Regression für komplexe Stichproben 37
 - Modell 38
 - Odds-Verhältnis 40
 - Optionen 41
 - Referenzkategorie 37
 - Statistik 38
 - Variablen speichern 40
 - zusätzliche Funktionen beim Befehl 41

M

- Martingale-Residuen
 - in Cox-Regression für komplexe Stichproben 54
- Maß für die Größe der Einheiten
 - beim Stichprobenassistenten 6
- Methode der Stichprobenziehung
 - beim Stichprobenassistenten 6
- Mittelwert
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
- Murthys Stichprobenmethode
 - beim Stichprobenassistenten 6

N

- Newton-Raphson-Methode
 - bei der ordinalen Regression für komplexe Stichproben 47

O

- Odds-Verhältnis
 - bei der logistischen Regression für komplexe Stichproben 40
 - bei der ordinalen Regression für komplexe Stichproben 46
 - in Kreuztabellen für komplexe Stichproben 25
- Ordinale Regression für komplexe Stichproben 43
 - Antwortwahrscheinlichkeiten 44
 - Modell 44
 - Odds-Verhältnis 46
 - Optionen 47
 - Statistik 45
 - Variablen speichern 47

P

- Parallelitätstest für Linien
 - bei der ordinalen Regression für komplexe Stichproben 45
- Parameterkonvergenz
 - bei der logistischen Regression für komplexe Stichproben 41

- Parameterkonvergenz (*Forts.*)
 - bei der ordinalen Regression für komplexe Stichproben 47
- Parameterschätzungen
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
 - in Cox-Regression für komplexe Stichproben 52
- Plandatei 2
- Polynomiale Kontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34
- PPS-Stichprobenziehung
 - beim Stichprobenassistenten 6
- Pseudo-R-Quadrat-Statistik
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45

Q

- Quadratwurzel aus dem Designeffekt
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Cox-Regression für komplexe Stichproben 52
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29

R

- R-Quadrat-Statistik
 - im allgemeinen linearen Modell für komplexe Stichproben 32
- Referenzkategorie
 - bei der logistischen Regression für komplexe Stichproben 37
 - im allgemeinen linearen Modell für komplexe Stichproben 34
- Relatives Risiko
 - in Kreuztabellen für komplexe Stichproben 25
- Residuen
 - im allgemeinen linearen Modell für komplexe Stichproben 34
 - in Kreuztabellen für komplexe Stichproben 25
- Risikodifferenz
 - in Kreuztabellen für komplexe Stichproben 25

S

- Sampfords Stichprobenmethode
 - beim Stichprobenassistenten 6
- Schichtung
 - beim Stichprobenassistenten 5
 - im Analysevorbereitungsassistenten 13
- Schoenfelds partielle Residuen
 - in Cox-Regression für komplexe Stichproben 54
- Schritthalbierungen
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47
- Scoreresiduen
 - in Cox-Regression für komplexe Stichproben 54
- Sequenzielle Bonferroni-Korrektur
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Sequenzielle Sidak-Korrektur
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Sequenzielle Stichprobenziehung
 - beim Stichprobenassistenten 6
- Sidak-Korrektur
 - in Cox-Regression für komplexe Stichproben 54
 - in komplexen Stichproben 33, 39, 46
- Spaltenprozenze
 - in Kreuztabellen für komplexe Stichproben 25
- Standardfehler
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29
- Stichprobe
 - komplexes Design 5
- Stichprobenanteil
 - beim Stichprobenassistenten 8
- Stichprobengewichtungen
 - beim Stichprobenassistenten 8
 - im Analysevorbereitungsassistenten 13
- Stichprobenplan 5
- Stichprobenschätzung
 - im Analysevorbereitungsassistenten 14
- Stichprobenumfang
 - beim Stichprobenassistenten 7, 8
- Summe
 - in "Deskriptive Statistiken für komplexe Stichproben" 23

- Systematische Stichprobenziehung
 - beim Stichprobenassistenten 6

T

- t-Test
 - bei der logistischen Regression für komplexe Stichproben 38
 - bei der ordinalen Regression für komplexe Stichproben 45
 - im allgemeinen linearen Modell für komplexe Stichproben 32
- Tabellenprozenze
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Kreuztabellen für komplexe Stichproben 25
- Teilgesamtheit
 - in Cox-Regression für komplexe Stichproben 52
- Test von proportionalen Hazards
 - in Cox-Regression für komplexe Stichproben 52
- Trennung
 - bei der logistischen Regression für komplexe Stichproben 41
 - bei der ordinalen Regression für komplexe Stichproben 47

U

- Umfang der Grundgesamtheit
 - beim Stichprobenassistenten 8
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29
- Ungewichtete Anzahl
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29

V

- Variationskoeffizient (COV)
 - in "Deskriptive Statistiken für komplexe Stichproben" 23
 - in "Häufigkeiten für komplexe Stichproben" 19
 - in Kreuztabellen für komplexe Stichproben 25
 - in Verhältnisse für komplexe Stichproben 29
- Verhältnisse für komplexe Stichproben
 - 29
 - fehlende Werte 30
 - Statistik 29

- Vorhergesagte Kategorien
 - bei der logistischen Regression für komplexe Stichproben 40
 - bei der ordinalen Regression für komplexe Stichproben 47
- Vorhergesagte Wahrscheinlichkeit
 - bei der logistischen Regression für komplexe Stichproben 40
 - bei der ordinalen Regression für komplexe Stichproben 47
- Vorhergesagte Werte
 - im allgemeinen linearen Modell für komplexe Stichproben 34

W

- Wiederholte Kontraste
 - im allgemeinen linearen Modell für komplexe Stichproben 34

Z

- Zeilenprozenze
 - in Kreuztabellen für komplexe Stichproben 25
- zeitabhängiger Prädiktor
 - in Cox-Regression für komplexe Stichproben 51

