

IBM SPSS Custom Tables 24

IBM

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 115 gelesen werden.

Produktinformation

Diese Ausgabe bezieht sich auf Version 24, Release 0, Modifikation 0 von IBM SPSS Statistics und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuauflage geändert wird.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs

IBM SPSS Custom Tables 24,

herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2016

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

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

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:

TSC Germany

Kst. 2877

Januar 2016

Inhaltsverzeichnis

Kapitel 1. Schnittstelle des Tabellenerstellungsprogramms 1

Schnittstelle des Tabellenerstellungsprogramms	1
Erstellen von Tabellen	1
Erstellen einer Tabelle	4
Stapeln von Variablen	4
Verschachteln von Variablen	4
Schichten	5
Einblenden und Ausblenden von Variablenamen bzw. -beschriftungen	6
Auswertungsstatistik	6
Kategorien und Gesamtsummen	12
Berechnete Kategorien	15
Tabellen für Variablen mit gemeinsamen Kategorien (Comperimeter-Tabellen)	17
Anpassen des Tabellenerstellungsprogramms	17
Benutzerdefinierte Tabellen: Registerkarte "Optionen"	17
Benutzerdefinierte Tabellen: Registerkarte "Titel"	19
Benutzerdefinierte Tabellen: Registerkarte "Teststatistiken"	19

Kapitel 2. Einfache Tabellen für kategoriale Variablen 23

Einfache Tabellen für kategoriale Variablen	23
Einzelne kategoriale Variable	23
Prozentsätze	24
Gesamtsummen	25
Kreuztabelle	25
Prozentsätze in Kreuztabellen	26
Steuern des Anzeigeformats	26
Randsummen	27
Sortieren und Ausschließen von Kategorien	28

Kapitel 3. Stapeln, Verschachteln und Schichten von kategorialen Variablen 31

Stapeln von kategorialen Variablen	31
Stapeln bei Kreuztabellen	31
Verschachteln von kategorialen Variablen	32
Unterdrücken von Variablenbeschriftungen	33
Verschachtelte Kreuztabellen	34
Schichten	35
Zwei gestapelte kategoriale Schichtvariablen	36
Zwei verschachtelte kategoriale Schichtvariablen	37

Kapitel 4. Gesamtsummen und Zwischenergebnisse für kategoriale Variablen 39

Einfache Gesamtsumme für eine Variable	39
Gesamtsumme entspricht der Summe aller angezeigten Werte	40
Anzeigeposition der Gesamtsummen	40
Gesamtsummen für verschachtelte Tabellen	41
Gesamtsummen für Schichtvariablen	42

Zwischenergebnisse	43
Zwischenergebnis entspricht der Summe der angezeigten Werte	43
Ausblenden der Kategorien für das Zwischenergebnis	44
Zwischenergebnisse für Schichtvariablen	45

Kapitel 5. Berechnete Kategorien für kategoriale Variablen 47

Einfache berechnete Kategorie	47
Ausblenden von Kategorien in einer berechneten Kategorie	48
Verweis auf Zwischenergebnisse in einer berechneten Kategorie	48
Einsatz von berechneten Kategorien zur Anzeige nicht abschließender Zwischenergebnisse	50

Kapitel 6. Tabellen für Variablen mit gemeinsamen Kategorien 53

Häufigkeitstabelle	53
Tabelle mit Prozentsätzen	54
Gesamtsummen und Kategorien	55
Verschachteln in Tabellen mit gemeinsamen Kategorien	56

Kapitel 7. Auswertungsstatistik 57

Quellenvariable für Auswertungsstatistik	57
Quelle von Auswertungsstatistiken für kategoriale Variablen	58
Quelle von Auswertungsstatistiken für metrische Variablen	58
Gestapelte Variablen	59
Angepasste Auswertungsstatistik für Gesamtsummen bei kategorialen Variablen	60
Anzeigen von Kategoriewerten	61

Kapitel 8. Auswerten von metrischen Variablen 63

Auswerten von metrischen Variablen	63
Gestapelte metrische Variablen	63
Mehrere Auswertungsstatistiken	64
Anzahl, gültige N und fehlende Werte	64
Unterschiedliche Auswertungen für unterschiedliche Variablen	66
Gruppierete Auswertungen in Kategorien	67
Mehrere Gruppierungsvariablen	67
Verschachteln von kategorialen in metrischen Variablen	68

Kapitel 9. Konfidenzintervalle 71

Kapitel 10. Teststatistiken 73

Teststatistiken	73
Tests auf Unabhängigkeit (Chi-Quadrat)	73

Auswirkungen von verschachtelten und gestapelten Variablen auf Tests auf Unabhängigkeit . . .	75
Vergleichen der Spaltenmittel	76
Auswirkungen von verschachtelten und gestapelten Variablen auf Spaltenmitteltests	78
Vergleichen der Spaltenanteile	79
Auswirkungen von verschachtelten und gestapelten Variablen auf Spaltenanteiletests	83
Anmerkung zu Gewichtungen und Mehrfachantwortsets	85
Kapitel 11. Mehrfachantwortsets	87
Häufigkeiten, Antworten, Prozentsätze und Gesamtwerte	87
Verwenden von Mehrfachantwortsets mit anderen Variablen	90
Statistikquellenvariablen und verfügbare Auswertungsstatistiken	90
Sets von kategorialen Variablen und doppelte Antworten	91
Signifikanztests bei Mehrfachantwortsets	92
Test auf Unabhängigkeit bei Mehrfachantwortsets	92
Vergleichen von Spaltenmitteln mit Mehrfachantwortsets	93

Kapitel 12. Fehlende Werte	95
Tabellen ohne fehlende Werte	95
Aufnehmen von fehlenden Werten in Tabellen.	96

Kapitel 13. Formatieren und Anpassen von Tabellen	99
Formatieren und Anpassen von Tabellen	99
Anzeigeformat für Auswertungsstatistiken	99
Anzeigen von Beschriftungen für Auswertungsstatistiken	100
Spaltenbreite	101
Anzeigewert für leere Zellen	102
Anzeigewert für fehlende Statistiken	102

Kapitel 14. Beispieldateien.	105
---	------------

Bemerkungen	115
Marken	116

Index	119
------------------------	------------

Kapitel 1. Schnittstelle des Tabellenerstellungsprogramms

Schnittstelle des Tabellenerstellungsprogramms

Mit dem Befehl "Benutzerdefinierte Tabellen" öffnen Sie eine Schnittstelle des Tabellenerstellungsprogramms, die mit Ziehen und Ablegen bedient werden kann und in der Sie eine Vorschau der Tabelle erhalten, während Sie Variablen und Optionen auswählen. Außerdem finden Sie hier einen Grad an Flexibilität, den ein normales Dialogfeld nicht bietet, beispielsweise die Möglichkeit, die Größe des Fensters und die Größe der einzelnen Fensterbereiche zu ändern.

Erstellen von Tabellen

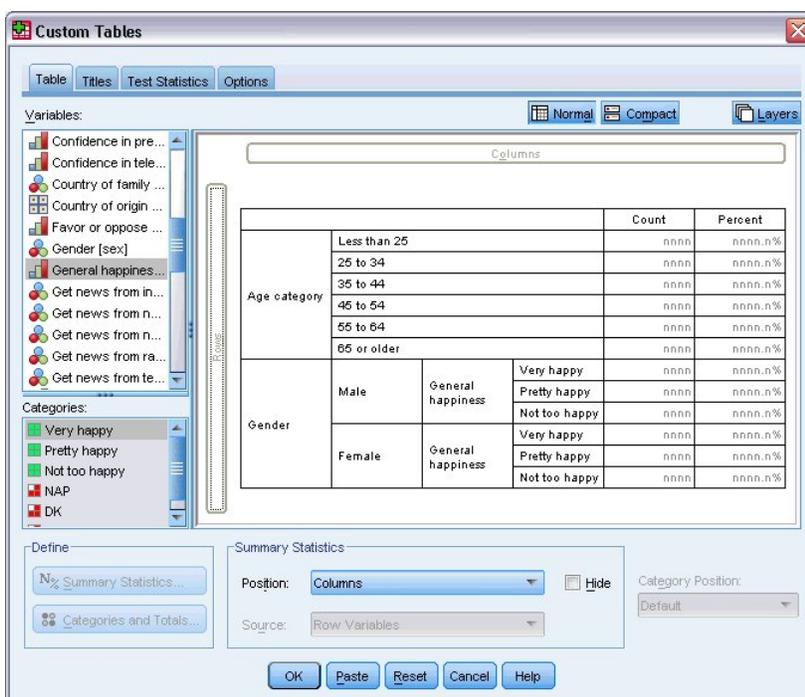


Abbildung 1. Dialogfeld "Benutzerdefinierte Tabellen", Registerkarte "Tabelle"

Auf der Registerkarte "Tabelle" des Tabellenerstellungsprogramms wählen Sie die Variablen und Auswertungsmaße aus, die in der Tabelle angezeigt werden sollen.

Liste "Variablen". Die Variablen aus der Datendatei werden links oben im Fenster angezeigt. Bei benutzerdefinierten Tabellen werden zwei Messniveaus unterschieden, nach denen Variablen behandelt werden:

Kategorial. Daten mit einer begrenzten Anzahl von eindeutigen Werten oder Kategorien (beispielsweise Geschlecht oder Religion). Kategoriale Variablen können Zeichenfolgevariablen (alphanumerisch) oder numerische Variablen sein, die zur Darstellung von Kategorien numerischen Code verwenden (beispielsweise 0 = männlich und 1 = weiblich). Auch als qualitative Daten bezeichnet. Kategoriale Variablen können **nominal** oder **ordinal** sein.

- *Nominal.* Eine Variable kann als nominal behandelt werden, wenn ihre Werte Kategorien darstellen, die sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Firmenabteilung, in der eine Person arbeitet. Beispiele für nominale Variablen sind Region, Postleitzahl oder Religionszugehörigkeit.

- *Ordinal*. Eine Variable kann als ordinal behandelt werden, wenn ihre Werte für Kategorien stehen, die eine natürliche Reihenfolge aufweisen (z. B. Grad der Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden). Ordinale Variablen treten beispielsweise bei Einstellungsmessungen (Zufriedenheit oder Vertrauen) und bei Präferenzbeurteilungen auf.

Skala. Daten, die auf einer Intervall- oder Verhältnisskala gemessen werden und bei denen die Datenwerte sowohl die Reihenfolge der Werte als auch die Distanz zwischen den Werten festlegen. So ist beispielsweise ein Gehalt von \$72.195 höher als ein Gehalt von \$52.398 und die Distanz zwischen den Werten beträgt \$19.797. Auch als quantitative oder stetige Daten bezeichnet.

Kategoriale Variablen definieren Kategorien (Zeilen, Spalten, Schichten) in der Tabelle. In der Standardeinstellung wird als Auswertungsstatistik die Anzahl (Anzahl der Fälle pro Kategorie) berechnet. In einer Standardtabelle für eine kategoriale Variable "Geschlecht" wird beispielsweise einfach die Anzahl der Männer und die Anzahl der Frauen aufgeführt.

Metrische Variablen werden normalerweise in den Kategorien von kategorialen Variablen ausgewertet. In der Standardeinstellung wird als Auswertungsstatistik der Mittelwert berechnet. So wird beispielsweise in einer Standardtabelle, in der das Einkommen nach Geschlecht kategorisiert ist, das durchschnittliche Einkommen von Männern und das durchschnittliche Einkommen von Frauen aufgeführt.

Es ist darüber hinaus möglich, metrische Variablen auszuwerten, ohne mithilfe von kategorialen Variablen Gruppen zu definieren. Dies ist in erster Linie zum **Stapeln** der Auswertungen von mehreren metrischen Variablen sinnvoll. Weitere Informationen finden Sie in „Stapeln von Variablen“ auf Seite 4.

Mehrfachantwortsets

In benutzerdefinierten Tabellen kann außerdem eine besondere Art von "Variable" verwendet werden, die als **Mehrfachantwortset** bezeichnet wird. Bei Mehrfachantwortsets handelt es sich nicht um Variablen im üblichen Sinn. Mehrfachantwortsets können nicht im Dateneditor angezeigt werden, sie werden von anderen Prozeduren nicht erkannt. Mehrfachantwortsets verwenden mehrere Variablen, um Antworten auf Fragen aufzuzeichnen, auf welche der Befragte mehr als eine Antwort geben kann. Sie werden wie kategoriale Variablen behandelt und bieten weitestgehend dieselben Möglichkeiten wie kategoriale Variablen. Weitere Informationen finden Sie in Kapitel 11, „Mehrfachantwortsets“, auf Seite 87.

Der Variablentyp ist durch ein Symbol neben der jeweiligen Variablen in der Variablenliste gekennzeichnet.

Tabelle 1. Messniveausymbole

	Numerisch	Zeichenfolge	Datum	Zeit
Metrisch (stetig)		entfällt		
Ordinal				
Nominal				

Tabelle 2. Symbole für Mehrfachantwortsets

Typ des Mehrfachantwortsets	Symbol
Mehrfachantwortset, mehrere Kategorien	

Tabelle 2. Symbole für Mehrfachantwortsets (Forts.)

Typ des Mehrfachantwortsets	Symbol
Mehrfachantwortset, mehrere Dichotomien	

Sie können das Messniveau einer Variablen ändern, indem Sie im Tabellenerstellungsprogramm mit der rechten Maustaste auf die Variable in der Variablenliste klicken und im Popup-Menü **Kategorial** oder **Metrisch** auswählen. In der Variablenansicht des Dateneditors kann das Messniveau einer Variablen dauerhaft geändert werden. Variablen, die als **nominal** oder **ordinal** definiert sind, werden bei benutzerdefinierten Tabellen als kategoriale Variablen behandelt.

Kategorien. Wenn Sie in der Variablenliste eine kategoriale Variable auswählen, werden die für diese Variable definierten Kategorien in der Liste "Kategorien" angezeigt. Die Kategorien werden außerdem im Erstellungsbereich angezeigt, wenn Sie die Variable in einer Tabelle verwenden. Wenn für die Variable keine Kategorien definiert wurden, werden in der Liste "Kategorien" und im Erstellungsbereich zwei Kategorieplatzhalter angezeigt: *Kategorie 1* und *Kategorie 2*.

Die definierten Kategorien, die im Tabellenerstellungsprogramm angezeigt werden, beruhen auf **Wertbeschriftungen**. Hierbei handelt es sich um aussagekräftige Beschriftungen, die verschiedenen Datenwerten zugeordnet sind (z. B. die beiden numerischen Werte 0 und 1 mit den jeweiligen Wertbeschriftungen *männlich* und *weiblich*). Wertbeschriftungen können in der Variablenansicht des Dateneditors oder mit dem Befehl "Variableneigenschaften definieren" aus dem Menü "Daten" definiert werden.

Erstellungsbereich. Tabellen werden durch Ziehen und Ablegen von Variablen auf den Zeilen und Spalten im Erstellungsbereich erstellt. Im Erstellungsbereich wird eine Vorschau der Tabelle angezeigt, die erstellt wird. Die Zellen im Erstellungsbereich enthalten keine tatsächlichen Datenwerte. Die Darstellung im Erstellungsbereich bietet jedoch eine relativ genaue Layoutansicht der resultierenden Tabelle. Bei kategorialen Variablen enthält die tatsächliche Tabelle möglicherweise mehr Kategorien als die Tabelle in der Vorschau, wenn sich in der Datendatei eindeutige Werte befinden, für die keine Wertbeschriftungen definiert wurden.

- In der **Normalansicht** werden alle Zeilen und Spalten angezeigt, die in der Tabelle enthalten sein werden, einschließlich der Zeilen bzw. Spalten für Auswertungsstatistiken und für die Kategorien der kategorialen Variablen.
- In der **Kompaktansicht** werden nur die Variablen angezeigt, die in der Tabelle enthalten sein werden, nicht jedoch die Vorschau der Zeilen und Spalten.

Grundregeln und Einschränkungen für das Erstellen von Tabellen

- Bei kategorialen Variablen beruhen Auswertungsstatistiken auf der innersten Variablen in der Quelldimension der Statistik.
- Die Standardquellendimension der Statistik (Zeile oder Spalte) für kategoriale Variablen beruht auf der Reihenfolge, in der Sie die Variablen in den Erstellungsbereich ziehen. Wenn Sie z. B. eine Variable zuerst in den Bereich für die Zeilen ziehen, wird die Zeilendimension als Standardquellendimension für die Statistik verwendet.
- Metrische Variablen können nur innerhalb der Kategorien der innersten Variablen in der Zeilen- oder der Spaltendimension ausgewertet werden. (Die metrische Variable kann auf einer beliebigen Ebene der Tabelle platziert werden, ihre Auswertung erfolgt jedoch auf der innersten Ebene.)
- Metrische Variablen können nicht innerhalb von anderen metrischen Variablen ausgewertet werden. Sie können Auswertungen von mehreren metrischen Variablen stapeln und metrische Variablen innerhalb der Kategorien von kategorialen Variablen auswerten. Es ist nicht möglich, metrische Variablen ineinander zu verschachteln oder eine metrische Variable in der Zeilendimension und eine andere in der Spaltendimension anzuordnen.

- Wenn eine Variable im aktiven Dataset mehr als 12.000 definierte Wertbeschriftungen enthält, können Sie das Tabellenerstellungsprogramm nicht zum Erstellen von Tabellen verwenden. Wenn Sie keine Variablen, die diese Beschränkung überschreiten, in Ihre Tabellen aufnehmen müssen, können Sie Variablensets definieren und anwenden, die diese Variablen ausschließen. Wenn Sie Variablen mit mehr als 12.000 definierten Wertbeschriftungen aufnehmen müssen, können Sie die betreffenden Tabellen mit der Befehlssyntax CTABLES generieren.

Erstellen einer Tabelle

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Ziehen Sie eine oder mehrere Variablen auf die Bereiche für Zeilen bzw. Spalten im Erstellungsbereich.
3. Klicken Sie auf **OK**, um die Tabelle zu erstellen.
4. Wählen Sie die Variable aus, indem Sie im Erstellungsbereich darauf klicken.
5. Ziehen Sie die Variable aus dem Erstellungsbereich heraus oder drücken Sie ENTF.
So ändern Sie das Messniveau einer Variablen:
6. Klicken Sie mit der rechten Maustaste in der Variablenliste auf die Variable (dies ist nur in der Variablenliste, nicht jedoch im Erstellungsbereich möglich).
7. Wählen Sie im Pop-up-Menü die Option **Kategorial** oder die Option **Metrisch** aus.

Stapeln von Variablen

Beim Stapeln werden separate Tabellen in dieselbe Anzeige eingefügt. So können beispielsweise Informationen über das *Geschlecht* und die *Alterskategorie* in unterschiedlichen Abschnitten derselben Tabelle angezeigt werden.

So stapeln Sie Variablen:

1. Wählen Sie in der Variablenliste alle zu stapelnden Variablen aus, ziehen Sie diese gemeinsam in den Erstellungsbereich und legen Sie sie im Bereich für die Zeilen oder die Spalten ab.
ODER
2. Ziehen Sie jede Variable einzeln und legen Sie diese oberhalb oder unterhalb von vorhandenen Variablen in den Zeilen bzw. links oder rechts von vorhandenen Variablen in den Spalten ab.

Tabelle 3. Gestapelte kategoriale Variablen.

Variablen	Kategorien	Auswertungsstatistik
Variable 1	Kategorie 1	123
	Kategorie 2	456
Variable 2	Kategorie 1	123
	Kategorie 2	456
	Kategorie 3	789

Weitere Informationen finden Sie in „Stapeln von kategorialen Variablen“ auf Seite 31.

Verschachteln von Variablen

Ähnlich wie bei Kreuztabellen kann mit Verschachtelung die Beziehung zwischen zwei kategorialen Variablen aufgezeigt werden. Bei der Verschachtelung wird eine der Variablen jedoch in derselben Dimension innerhalb der anderen Variablen verschachtelt. So können Sie beispielsweise *Geschlecht* in *Alterskategorie* in der Zeilendimension verschachteln, um die Anzahl der Männer und Frauen in jeder Alterskategorie darzustellen.

Es ist außerdem möglich, metrische Variablen in kategorialen Variablen zu verschachteln. So können Sie beispielsweise *Einkommen* in *Geschlecht* verschachteln, sodass das mittlere Einkommen (oder der Median bzw. ein anderes Auswertungsmaß des Einkommens) für Frauen und Männer getrennt aufgeführt wird.

So verschachteln Sie Variablen:

1. Ziehen Sie eine kategoriale Variable in den Erstellungsbereich und legen Sie sie im Bereich für die Zeilen oder die Spalten ab.
2. Ziehen Sie eine kategoriale oder eine metrische Variable in den Erstellungsbereich und legen Sie sie links oder rechts von der kategorialen Zeilenvariablen bzw. oberhalb oder unterhalb der kategorialen Spaltenvariablen ab.

Table 4. Verschachtelte kategoriale Variablen.

Variable 1	Variable 2	Auswertungsstatistik
Kategorie 1	Kategorie 1	12
	Kategorie 2	34
	Kategorie 3	56
Kategorie 2	Kategorie 1	12
	Kategorie 2	34
	Kategorie 3	56

Weitere Informationen finden Sie in „Verschachteln von kategorialen Variablen“ auf Seite 32.

Hinweis: Bei benutzerdefinierten Tabellen wird die Verarbeitung geschichteter aufgeteilter Dateien nicht berücksichtigt. Um dasselbe Ergebnis wie bei geschichteten aufgeteilten Dateien zu erzielen, müssen Sie die Dateiteilungsvariablen in den äußersten Verschachtelungsebenen der Datei platzieren.

Schichten

Mit Schichten können Tabellen um eine Tiefendimension erweitert werden, um dreidimensionale "Würfel" zu erstellen. Schichten ähneln dem Stapeln und Verschachteln. Der wichtigste Unterschied besteht darin, dass stets nur eine Schichtenkategorie sichtbar ist. Wenn Sie beispielsweise *Alterskategorie* als Zeilenvariable und *Geschlecht* als Schichtvariable verwenden, erhalten Sie eine Tabelle, in der die Informationen für Männer und Frauen in unterschiedlichen Schichten dargestellt werden.

So erstellen Sie Schichten:

1. Klicken Sie auf der Registerkarte "Tabelle" des Tabellenerstellungsprogramms auf **Schichten**, um die Liste "Schichten" anzuzeigen.
2. Ziehen Sie die zum Definieren der Schichten vorgesehenen metrischen oder kategorialen Variablen in die Liste "Schichten" und legen Sie diese dort ab.

Metrische und kategoriale Variablen können in der Liste "Schichten" nicht kombiniert werden. Alle Variablen müssen denselben Typ aufweisen. Mehrfachantwortsets werden in der Liste "Schichten" wie kategoriale Variablen behandelt. Metrische Variablen in den Schichten werden immer gestapelt.

Wenn mehrere kategoriale Schichtvariablen vorliegen, können Schichten gestapelt oder verschachtelt werden.

- **Jede Kategorie als Schicht zeigen** entspricht einer Stapelung. Für jede Kategorie jeder Schichtvariablen wird eine eigene Schicht angezeigt. Die Gesamtzahl der Schichten ergibt sich einfach aus der *Summe* der Anzahl Kategorien jeder Schichtvariablen. Wenn beispielsweise drei Schichtvariablen vorhanden sind, von denen jede drei Kategorien besitzt, enthält die Tabelle neun Schichten.
- **Jede Kombination von Kategorien als Schicht zeigen** entspricht der Verschachtelung bzw. dem Erstellen von Kreuztabellen für Schichten. Die Gesamtzahl der Schichten ergibt sich aus dem *Produkt* der

Anzahl Kategorien jeder Schichtvariablen. Wenn beispielsweise drei Variablen vorhanden sind, von denen jede drei Kategorien besitzt, enthält die Tabelle 27 Schichten.

Einblenden und Ausblenden von Variablennamen bzw. -beschriftungen

Für die Anzeige von Variablennamen und -beschriftungen sind die folgenden Optionen verfügbar:

- Nur Variablenbeschriftungen anzeigen. Bei Variablen, für die keine Variablenbeschriftung definiert ist, wird der Variablenname angezeigt. Dies ist die Standardeinstellung.
- Nur Variablennamen anzeigen.
- Sowohl Variablenbeschriftungen als auch Variablennamen anzeigen.
- Weder Variablennamen noch Variablenbeschriftungen anzeigen. Die Spalte bzw. Zeile mit der Variablenbeschriftung oder dem Variablennamen wird zwar weiterhin in der Tabellenvorschau im Erstellungsbereich angezeigt, in der tatsächlichen Tabelle ist diese Spalte bzw. Zeile dann jedoch ausgeblendet.

So zeigen Sie Variablenbeschriftungen und -namen an oder blenden sie aus:

1. Klicken Sie mit der rechten Maustaste auf die Tabellenvorschau im Erstellungsbereich.
2. Wählen Sie im Popup-Menü die Option **Variablenbeschriftung anzeigen** oder die Option **Variablenname anzeigen** aus, um zwischen ein- und ausgeblendeten Beschriftungen und Namen zu wechseln. Ein Häkchen neben einer Option zeigt an, dass die entsprechenden Angaben angezeigt werden.

Auswertungsstatistik

Das Dialogfeld "Auswertungsstatistik" bietet die folgenden Funktionen:

- Hinzufügen und Entfernen von Auswertungsstatistiken zu bzw. aus einer Tabelle
- Ändern der Beschriftung von Statistiken
- Ändern der Reihenfolge der Statistiken
- Ändern des Formats einer Statistik, einschließlich der Anzahl Dezimalstellen

Es hängt vom Messniveau der jeweiligen Quellvariablen ab, welche Auswertungsstatistiken (und andere Optionen) verfügbar sind, wie oben im Dialogfeld dargestellt. Die Quelle der Auswertungsstatistiken (die Variable, auf der die Auswertungsstatistiken beruhen) wird durch die folgenden Faktoren bestimmt:

- **Messniveau.** Wenn eine Tabelle (oder ein Tabellenabschnitt in einer gestapelten Tabelle) eine metrische Variable enthält, beruhen Auswertungsstatistiken auf der metrischen Variablen.
- **Reihenfolge der Variablenauswahl.** Die Standardquellendimension der Statistik (Zeile oder Spalte) für kategoriale Variablen beruht auf der Reihenfolge, in der Sie die Variablen in den Erstellungsbereich ziehen. Wenn Sie z. B. eine Variable zuerst in den Bereich für die Zeilen ziehen, wird die Zeilendimension als Standardquellendimension für die Statistik verwendet.
- **Verschachtelung.** Bei kategorialen Variablen beruhen Auswertungsstatistiken auf der innersten Variablen in der Quellendimension der Statistik.

Eine gestapelte Tabelle kann mehrere Quellvariablen für Auswertungsstatistiken aufweisen (sowohl metrisch als auch kategorial); jeder Tabellenabschnitt besitzt jedoch nur eine Quelle für die Auswertungsstatistik.

So ändern Sie die Quellendimension der Auswertungsstatistik:

1. Wählen Sie auf der Registerkarte "Tabelle" im Gruppenfeld "Auswertungsstatistik" in der Dropdown-Liste **Quelle** die Dimension (Zeilen, Spalten oder Schichten) aus.

So legen Sie die in einer Tabelle angezeigte Auswertungsstatistik fest:

1. Wählen Sie die Quellvariable für die Auswertungsstatistik aus, indem Sie im Erstellungsbereich der Registerkarte "Tabelle" darauf klicken.

2. Klicken Sie auf der Registerkarte "Tabelle" im Gruppenfeld "Definieren" auf **Auswertungsstatistik**.
ODER
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf die Quellenvariable für die Auswertungsstatistik und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus.
4. Wählen Sie die Auswertungsstatistik aus, die Sie in die Tabelle aufnehmen möchten. Mit der Pfeilschaltfläche können Sie ausgewählte Statistiken aus der Liste "Statistik" in die Liste "Anzeigen" übertragen. Sie können die Statistiken auch aus der Liste "Statistik" in die Liste "Anzeigen" ziehen und dort ablegen.
5. Klicken Sie auf die nach oben oder unten weisende Pfeilschaltfläche, um die Anzeigeposition der aktuell ausgewählten Auswertungsstatistik zu ändern.
6. Wählen Sie aus der Dropdown-Liste "Format" ein Anzeigeformat für die ausgewählte Auswertungsstatistik aus.
7. Geben Sie in der Zelle "Dezimalstellen" die Anzahl der anzuzeigenden Dezimalstellen für die ausgewählte Auswertungsstatistik ein.
8. Klicken Sie auf **Der Auswahl zuweisen**, um die ausgewählten Auswertungsstatistiken für die aktuell ausgewählten Variablen im Erstellungsbereich in die Tabelle aufzunehmen.
9. Klicken Sie auf **Allen zuweisen**, um die ausgewählten Auswertungsstatistiken für alle gestapelten Variablen desselben Typs im Erstellungsbereich in die Tabelle aufzunehmen.

Hinweis: **Allen zuweisen** unterscheidet sich nur bei gestapelten Variablen gleichen Typs, die sich bereits im Erstellungsbereich befinden, von **Der Auswahl zuweisen**. In beiden Fällen werden die ausgewählten Auswertungsstatistiken automatisch auf alle weiteren gestapelten Variablen gleichen Typs angewendet, die Sie der Tabelle hinzufügen.

Auswertungsstatistiken für kategoriale Variablen

Als grundlegende Statistiken für kategoriale Variablen sind Häufigkeiten und Prozentsätze verfügbar. Zusätzlich können angepasste Auswertungsstatistiken für Gesamtsummen und Zwischenergebnisse festgelegt werden. Diese angepassten Auswertungsstatistiken umfassen Lagemaße (wie Mittelwert und Median) und Streuungsmaße (wie Standardabweichung), die sich für einige ordinale kategoriale Variablen eignen. Weitere Informationen finden Sie in „Angepasste Auswertungsstatistik für Gesamtsummen bei kategorialen Variablen“ auf Seite 11.

Häufigkeiten. Anzahl der Fälle in jeder Zelle einer Tabelle bzw. Anzahl der Antworten bei Mehrfachantwortsets. Wenn die Gewichtung aktiviert ist, entspricht dieser Wert der gewichteten Anzahl.

- Wenn die Gewichtung aktiviert ist, entspricht der Wert der gewichteten Anzahl.
- Die gewichtete Anzahl ist bei der Gewichtung globaler Datensets (**Daten > Fälle gewichten**) und bei der Gewichtung mit einer Gewichtungsvariablen für eine effektive Basis, die über die Registerkarte **Optionen** des Dialogfelds **Benutzerdefinierte Tabellen** angegeben wird, identisch.

Ungewichtete Anzahl. Ungewichtete Anzahl von Fällen in jeder Zelle der Tabelle. Diese unterscheidet sich von der Anzahl nur, wenn eine Gewichtung verwendet wird.

Gewichtete Anzahl. Die gewichtete Anzahl, die in Gewichtungsberechnungen für eine effektive Basis verwendet wird. Wenn Sie keine Gewichtungsvariable für eine effektive Basis (Registerkarte **Optionen**) verwenden, ist die gewichtete Anzahl mit der Anzahl identisch.

Spaltenprozente. Prozentsätze in jeder Spalte. Die Summe der Prozentsätze in jeder Spalte einer Untertabelle beträgt 100 % (bei einfachen Prozentsätzen). Spaltenprozente sind in der Regel nur dann sinnvoll, wenn eine kategoriale *Zeilenvariable* vorhanden ist.

Zeilenprozente. Prozentsätze in jeder Zeile. Die Summe der Prozentsätze in jeder Zeile einer Untertabelle beträgt 100 % (bei einfachen Prozentsätzen). Zeilenprozente sind in der Regel nur dann sinnvoll, wenn eine kategoriale *Spaltenvariable* vorhanden ist.

Zeilen- und Spaltenprozentage in Schichten. Die Summe der Zeilen- oder Spaltenprozentage über alle Untertabellen in einer verschachtelten Tabelle beträgt 100 % (bei einfachen Prozentsätzen). Wenn die Tabelle Schichten enthält, ergibt die Summe der Zeilen- oder Spaltenprozentage über alle verschachtelten Untertabellen jeder Schicht 100 %.

Schichtenprozentage. Prozentsätze in jeder Schicht. Bei einfachen Prozentsätzen ergibt die Summe der Prozentsätze für Zellen in der aktuell sichtbaren Schicht 100 %. Wenn keine Schichtvariablen vorhanden sind, entspricht dies den Tabellenprozentagen.

Tabellenprozentage. Die Prozentsätze für jede Zelle beziehen sich auf die gesamte Tabelle. Alle Prozentsätze für Zellen beruhen auf derselben Gesamtanzahl von Fällen; die Summe der Prozentsätze über die gesamte Tabelle beträgt 100 % (bei einfachen Prozentsätzen).

Untertabellenprozentage. Die Prozentsätze für jede Zelle beziehen sich auf die Untertabelle. Alle Prozentsätze für Zellen in der Untertabelle beruhen auf derselben Gesamtanzahl von Fällen; die Summe der Prozentsätze über die gesamte Untertabelle beträgt 100 %. In verschachtelter Tabelle werden die Untertabellen durch die Variable definiert, die der innersten Verschachtelungsebene vorangeht. So werden beispielsweise die Untertabellen in einer Tabelle mit *Familienstand* innerhalb von *Geschlecht* innerhalb von *Alterskategorie* durch die Variable *Geschlecht* definiert.

Konfidenzintervalle

- Untere und obere Konfidenzgrenzen sind für Häufigkeiten, Prozentsätze, Mittelwerte, Mediane, Perzentile und Summen verfügbar.
- Die Textzeichenfolge "&[Konfidenzniveau]" in der Beschriftung enthält das Konfidenzniveau in der Spaltenbeschriftung in der Tabelle.
- Der Standardfehler ist für Häufigkeiten, Prozentsätze, Mittelwerte und Summen verfügbar.
- Konfidenzintervalle und Standardfehler sind für Mehrfachantwortsets nicht verfügbar.

Niveau

Das Konfidenzniveau für Konfidenzintervalle, ausgedrückt in Prozent. Der Wert muss größer als 0 und kleiner als 100 sein.

Mehrfachantwortsets

Die Prozentsätze von Mehrfachantwortsets können sich auf Fälle, Antworten oder Häufigkeiten beziehen. Weitere Informationen finden Sie in „Auswertungsstatistiken für Mehrfachantwortsets“ auf Seite 9.

Gestapelte Tabellen

Bei der Berechnung der Prozentsätze wird jeder von einer gestapelten Variablen definierte Tabellenabschnitt als eigene Tabelle behandelt. Die Summe von Zeilen- und Spaltenprozentagen in Schichten sowie Tabellenprozentagen in jedem gestapelten Tabellenabschnitt beträgt 100 % (bei einfachen Prozentsätzen). Bei anderen Prozentberechnungen basieren die Prozentsätze auf den Fällen in jedem gestapelten Tabellenabschnitt.

Prozentbasis

Abhängig von der Behandlung fehlender Werte in der Berechnungsgrundlage können Prozentsätze auf drei verschiedenen Arten berechnet werden:

Einfacher Prozentsatz. Die Prozentsätze beziehen sich auf die Anzahl der Fälle, die in der Tabelle verwendet werden. Die Summe der Prozentsätze beträgt stets 100 %. Wenn eine Kategorie aus der Tabelle ausgeschlossen ist, werden die Fälle dieser Kategorie bei der Berechnung nicht berücksichtigt. Fälle mit systemdefiniert fehlenden Werten werden immer aus der Berechnungsgrundlage ausgeschlossen. Fälle mit benutzerdefiniert fehlenden Werten werden bei der Berechnung nicht berücksichtigt, wenn benutzerdefiniert fehlende Kategorien aus der Tabelle ausgeschlossen sind (Standardeinstellung). Sie werden be-

rücksichtigt, wenn benutzerdefiniert fehlende Kategorien in der Tabelle enthalten sind. Alle Prozentsätze ohne *Gültige N* oder *Gesamtanzahl* in der Bezeichnung sind einfache Prozentsätze.

Prozentsätze auf Grundlage der Gesamtanzahl. Der Prozentbasis werden die Fälle mit systemdefiniert fehlenden und benutzerdefiniert fehlenden Werten hinzugefügt. Die Summe der Prozentsätze kann weniger als 100 % betragen.

Prozentsätze auf Grundlage der gültigen N. Fälle mit benutzerdefiniert fehlenden Werten werden aus der einfachen Prozentbasis entfernt, auch wenn benutzerdefiniert fehlende Kategorien in der Tabelle enthalten sind.

Hinweis: Fälle in manuell ausgeschlossenen Kategorien, die keine benutzerdefiniert fehlenden Kategorien sind, werden bei der Berechnung nie berücksichtigt.

Auswertungsstatistiken für Mehrfachantwortsets

Für Mehrfachantwortsets sind die folgenden zusätzlichen Auswertungsstatistiken verfügbar.

Antworten als Spalten%/Zeilen%/Schichten%. Die Prozentsätze basieren auf Antworten.

Antworten als Spalten%/Zeilen%/Schichten% (Basis: Anzahl). Der Zähler enthält die Antworten, der Nenner die Gesamtanzahl.

Anzahl als Spalten%/Zeilen%/Schichten% (Basis: Antworten). Der Zähler enthält die Anzahl, der Nenner die Gesamtanzahl der Antworten.

Antworten als Spalten%/Zeilen% in Schicht. Prozentsatz für mehrere Untertabellen. Die Prozentsätze basieren auf Antworten.

Antworten als Spalten%/Zeilen% in Schicht (Basis: Anzahl). Prozentsätze für mehrere Untertabellen. Der Zähler enthält die Antworten, der Nenner die Gesamtanzahl.

Antworten als Spalten%/Zeilen% in Schicht (Basis: Antworten). Prozentsätze für mehrere Untertabellen. Der Zähler enthält die Anzahl, der Nenner die Gesamtanzahl der Antworten.

Antworten. Anzahl der Antworten.

Antworten als Untertabellen%/Tabellen%. Die Prozentsätze basieren auf Antworten.

Antworten als Untertabellen%/Tabellen% (Basis: Anzahl). Der Zähler enthält die Antworten, der Nenner die Gesamtanzahl.

Anzahl als Untertabellen%/Tabellen% (Basis: Antworten). Der Zähler enthält die Anzahl, der Nenner die Gesamtanzahl der Antworten.

Auswertungsstatistiken für metrische Variablen und angepasste Gesamtsummen für kategoriale Variablen

Zusätzlich zu den für kategoriale Variablen verfügbaren Häufigkeiten und Prozentsätzen sind für metrische Variablen die folgenden Auswertungsstatistiken verfügbar, die auch als angepasste Auswertungen für Gesamtsummen und Zwischenergebnisse bei kategorialen Variablen verwendet werden können. Diese Auswertungsstatistiken sind für Mehrfachantwortsets oder (alphanumerische) Zeichenfolgevariablen nicht verfügbar.

Mittelwert. Arithmetisches Mittel; die Summe geteilt durch die Anzahl der Fälle.

Median. Wert, über und unter dem jeweils die Hälfte der Fälle liegt; 50. Perzentil.

Modalwert. Häufigster Wert. Wenn zwei oder mehr Werte gleich häufig auftreten, wird der kleinste Wert angezeigt.

Minimum. Kleinster (niedrigster) Wert.

Maximum. Größter (höchster) Wert.

Fehlend. Anzahl der fehlenden Werte (sowohl benutzerdefiniert fehlende als auch systemdefiniert fehlende Werte).

Perzentil. Sie können das 5., 25., 75., 95. und/oder 99. Perzentil abrufen.

Bereich. Differenz zwischen größtem und kleinstem Wert.

Standardabweichung Ein Maß für die Streuung um den Mittelwert. In einer Normalverteilung liegen 68 % der Fälle innerhalb von einer Standardabweichung des Mittelwerts und 95 % der Fälle innerhalb von zwei Standardabweichungen. Wenn beispielsweise das durchschnittliche Alter 45 und die Standardabweichung 10 beträgt, liegen bei einer Normalverteilung 95 % der Fälle im Bereich zwischen 25 und 65 Jahren (Quadratwurzel der Varianz).

Summe. Die Summe der Werte.

Prozent der Summe. Prozentsätze bezogen auf Summen. Verfügbar für Zeilen und Spalten (von Untertabellen), ganze Zeilen und Spalten (über mehrere Untertabellen), Schichten, Untertabellen und ganze Tabellen.

Gesamtanzahl. Anzahl der nicht fehlenden, benutzerdefiniert fehlenden und systemdefiniert fehlenden Werte. Fälle in manuell ausgeschlossenen Kategorien, die nicht zu den benutzerdefiniert fehlenden Kategorien gehören, werden hierbei nicht berücksichtigt.

Korrigierte Gesamtanzahl. Die korrigierte Gesamtanzahl wird in Gewichtungsberechnungen für eine effektive Basis verwendet. Wenn Sie keine Gewichtungsvariable für eine effektive Basis (Registerkarte **Optionen**) verwenden, ist die korrigierte Gesamtanzahl mit der Gesamtanzahl identisch. Diese Statistik ist für Mehrfachantwortsets nicht verfügbar.

Gültige Anzahl. Anzahl der nicht fehlenden Werte. Fälle in manuell ausgeschlossenen Kategorien, die nicht zu den benutzerdefiniert fehlenden Kategorien gehören, werden hierbei nicht berücksichtigt.

Korrigierte gültige Anzahl. Die korrigierte gültige Anzahl wird in Gewichtungsberechnungen für eine effektive Basis verwendet. Wenn Sie keine Gewichtungsvariable für eine effektive Basis (Registerkarte **Optionen**) verwenden, ist die korrigierte gültige Anzahl mit der gültigen Anzahl identisch. Diese Statistik ist für Mehrfachantwortsets nicht verfügbar.

Varianz. Ein Maß der Streuung um den Mittelwert, gleich der Summe der quadrierten Abweichungen vom Mittelwert geteilt durch eins weniger als die Anzahl der Fälle. Die Einheit der Varianz entspricht der quadrierten Einheit der eigentlichen Variablen (quadrierte Standardabweichung).

Konfidenzintervalle

- Untere und obere Konfidenzgrenzen sind für Häufigkeiten, Prozentsätze, Mittelwerte, Mediane, Perzentile und Summen verfügbar.
- Die Textzeichenfolge "&[Konfidenzniveau]" in der Beschriftung enthält das Konfidenzniveau in der Spaltenbeschriftung in der Tabelle.
- Der Standardfehler ist für Häufigkeiten, Prozentsätze, Mittelwerte und Summen verfügbar.
- Konfidenzintervalle und Standardfehler sind für Mehrfachantwortsets nicht verfügbar.

Niveau

Das Konfidenzniveau für Konfidenzintervalle, ausgedrückt in Prozent. Der Wert muss größer als 0 und kleiner als 100 sein.

Gestapelte Tabellen

Jeder von einer gestapelten Variablen definierte Tabellenabschnitt wird als eigene Tabelle behandelt und die Auswertungsstatistiken werden dementsprechend berechnet.

Angepasste Auswertungsstatistik für Gesamtsummen bei kategorialen Variablen

Tabellen mit kategorialen Variablen, die Gesamtsummen oder Zwischenergebnisse enthalten, können andere Auswertungsstatistiken als die Zusammenfassungen enthalten, die für jede Kategorie angezeigt werden. So können beispielsweise Häufigkeiten und Spaltenprozente für eine ordinale kategoriale Zeilenvariable und der Median für die Gesamtsummen angezeigt werden.

So erstellen Sie eine Tabelle für eine kategoriale Variable mit einer angepassten Gesamtsumme:

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

Das Tabellenerstellungsprogramm wird geöffnet.

2. Ziehen Sie eine kategoriale Variable in den Erstellungsbereich und legen Sie sie im Bereich für die Zeilen oder die Spalten ab.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf die Variable und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
4. Klicken Sie auf das Kontrollkästchen **Gesamt** (aktivieren Sie dieses) und klicken Sie anschließend auf **Zuweisen**.
5. Klicken Sie erneut mit der rechten Maustaste im Erstellungsbereich auf die Variable und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus.
6. Klicken Sie auf das Kontrollkästchen **Angepasste Auswertungsstatistik für Gesamtsummen und Zwischenergebnisse** (aktivieren Sie dieses) und wählen Sie anschließend die gewünschte angepasste Auswertungsstatistik.

In der Standardeinstellung werden alle Auswertungsstatistiken einschließlich der angepassten in der entgegengesetzten Dimension der kategorialen Variablen dargestellt. Wenn beispielsweise eine kategoriale Zeilenvariable vorhanden ist, definieren die Auswertungsstatistiken wie im folgenden Beispiel die Spalten der Tabelle:

Tabelle 5. Kategorien von ordinalen Variablen in Zeilen, Auswertungsstatistikanzahl und Mittelwert in Spalten.

Variablen	Kategorien	Anzahl	Mittelwert
Variable 1	1 Zustimmung	196	2,29
	2 Neutral	936	
	3 Ablehnung	744	
	Gesamtergebnis	1876	

So zeigen Sie Auswertungsstatistiken in derselben Dimension wie die kategoriale Variable an:

7. Wählen Sie im Tabellenerstellungsprogramm auf der Registerkarte "Tabelle" im Gruppenfeld "Auswertungsstatistik" in der Dropdown-Liste "Position" die Dimension.

Wenn die kategoriale Variable beispielsweise in den Zeilen angezeigt wird, wählen Sie in der Dropdown-Liste die Option **Zeilen**.

Anzeigeformate für Auswertungsstatistiken

Für die Anzeigeformate sind die folgenden Optionen verfügbar:

nnnn. Einfach numerisch.

nnnn%. Wert mit nachgestelltem Prozentzeichen.

Auto. Definiertes Anzeigeformat für die Variable (einschließlich der Dezimalstellen).

N=nnnn. Zeigt den Wert mit vorangestelltem *N=* an. Dies ist bei Häufigkeiten, *Gültige N* und *Gesamtanzahl* in Tabellen sinnvoll, in denen die Beschriftungen für Auswertungsstatistiken nicht angezeigt werden.

(nnnn). Alle Werte werden in Klammern angezeigt.

(nnnn)(neg. Wert). Nur negative Werte werden in Klammern angezeigt.

(nnnn%). Alle Werte werden in Klammern und mit nachgestelltem Prozentzeichen angezeigt.

n,nnn.n. Kommaformat. Unabhängig von den regionalen Einstellungen bzw. Ländereinstellungen von Windows wird das Komma als Tausender- und der Punkt als Dezimaltrennzeichen verwendet.

n.nnn,n. Punktformat. Unabhängig von den regionalen Einstellungen bzw. Ländereinstellungen von Windows wird der Punkt als Tausender- und das Komma als Dezimaltrennzeichen verwendet.

\$n,nnn.n. Dollarformat. Unabhängig von den regionalen Einstellungen bzw. Ländereinstellungen von Windows wird das Dollarzeichen dem Wert vorangestellt und das Komma als Tausender- sowie der Punkt als Dezimaltrennzeichen verwendet.

CCA, CCB, CCC, CCD, CCE. Währungsformate. In der Liste werden die aktuell definierten Formate jeder Währung angezeigt. Diese Formate werden im Dialogfeld "Optionen" (Menü "Bearbeiten", Befehl "Optionen") auf der Registerkarte "Währung" definiert.

Allgemeine Regeln und Einschränkungen

- Außer beim Format "Auto" richtet sich die Anzahl der Dezimalstellen nach der Einstellung in der Spalte "Dezimalstellen".
- Außer bei den Formaten "Komma", "Dollar" und "Punkt" wird als Dezimaltrennzeichen das in der Systemsteuerung von Windows unter "Ländereinstellungen" bzw. "Regionale Einstellungen" aktuell festgelegte Zeichen verwendet.
- Bei den Formaten "Komma", "Dollar" und "Punkt" wird als Tausendertrennzeichen ein Komma oder ein Punkt verwendet. Beim Erstellen einer Tabelle ist jedoch kein Anzeigeformat verfügbar, mit dem das Tausendertrennzeichen aus den aktuellen regionalen Einstellungen bzw. Ländereinstellungen von Windows (Systemsteuerung, Option "Ländereinstellungen" bzw. "Regionale Einstellungen") übernommen wird.

Kategorien und Gesamtsummen

Das Dialogfeld "Kategorien und Gesamtsummen" bietet die folgenden Funktionen:

- Umordnen und Ausschließen von Kategorien.
- Einfügen von Gesamtsummen und Zwischenergebnissen.
- Einfügen von berechneten Kategorien.
- Ein- oder Ausschließen von leeren Kategorien.
- Ein- oder Ausschließen von als fehlend definierten Kategorien.
- Ein- oder Ausschließen von Kategorien, für die keine Wertbeschriftungen definiert wurden.

- Dieses Dialogfeld ist nur für kategoriale Variablen und Mehrfachantwortsets verfügbar. Es ist nicht für metrische Variablen verfügbar.
- Bei mehreren ausgewählten Variablen mit unterschiedlichen Kategorien können keine Zwischenergebnisse und berechneten Kategorien eingefügt, Kategorien ausgeschlossen oder Kategorien manuell umgeordnet werden. Dies ist aber nur dann der Fall, wenn Sie in der Vorschau im Erstellungsbereich mehrere Variablen auswählen und anschließend das Dialogfeld für alle ausgewählten Variablen öffnen. Diese Vorgänge können jedoch für jede Variable einzeln durchgeführt werden.
- Für Variablen ohne Wertbeschriftungen können nur Kategorien sortiert und Gesamtsummen eingefügt werden.

So öffnen Sie das Dialogfeld "Kategorien und Gesamtsummen":

1. Ziehen Sie eine kategoriale Variable bzw. ein Mehrfachantwortset in den Erstellungsbereich und legen Sie diese bzw. dieses dort ab.
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf die Variable und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
ODER
3. Wählen Sie die Variable im Erstellungsbereich aus, indem Sie darauf klicken, und klicken Sie anschließend auf der Registerkarte "Tabelle" im Gruppenfeld "Definieren" auf **Kategorien und Gesamtsummen**.
Sie können außerdem im Erstellungsbereich mehrere kategoriale Variablen in derselben Dimension auswählen:
4. Klicken Sie bei gedrückter Steuertaste im Erstellungsbereich auf die Variablen.
ODER
5. Klicken Sie im Erstellungsbereich auf einen Bereich außerhalb der Tabellenvorschau. Wählen Sie anschließend durch Klicken und Ziehen einen Bereich aus, der die gewünschten Variablen enthält.
ODER
6. Klicken Sie mit der rechten Maustaste auf eine Variable in einer Dimension und wählen Sie die Option **Alle [Dimension]variablen auswählen**, um alle Variablen in dieser Dimension auszuwählen.

Umordnen von Kategorien

So ordnen Sie Kategorien manuell um:

1. Wählen Sie in der Liste eine Kategorie aus, indem Sie darauf klicken.
2. Klicken Sie auf die nach oben oder die nach unten weisende Pfeilschaltfläche, um die Kategorie in der Liste nach oben oder unten zu verschieben.
ODER
3. Klicken Sie auf die Spalte "Wert(e)" der Kategorie und verändern Sie durch Ziehen und Ablegen deren Position.

So schließen Sie Kategorien aus:

1. Wählen Sie in der Liste eine Kategorie aus, indem Sie darauf klicken.
2. Klicken Sie auf die Pfeilschaltfläche neben der Liste "Ausschließen".
ODER
3. Klicken Sie auf die Spalte "Wert(e)" der Kategorie und ziehen Sie sie aus der Liste.

Wenn Sie eine Kategorie ausschließen, werden alle Kategorien, für die keine Wertbeschriftungen definiert sind, ebenfalls ausgeschlossen.

So sortieren Sie Kategorien:

Kategorien können nach Datenwerten, Wertbeschriftungen, Zellenhäufigkeit oder Auswertungsstatistik in absteigender oder aufsteigender Reihenfolge sortiert werden.

1. Wählen Sie im Gruppenfeld "Kategorien sortieren" in der Dropdown-Liste "Nach" das gewünschte Sortierkriterium aus: Wert, Beschriftung, Anzahl oder Auswertungsstatistik (z. B. Mittelwert, Median, Modalwert). Welche Auswertungsstatistik für die Sortierung verfügbar ist, hängt davon ab, welche Auswertungsstatistik Sie für die Anzeige in der Tabelle ausgewählt haben.
2. Wählen Sie in der Dropdown-Liste "Reihenfolge" die Sortierreihenfolge (aufsteigend, absteigend) aus.

Wenn Sie Kategorien ausgeschlossen haben, ist die Funktion zum Sortieren von Kategorien nicht verfügbar.

Zwischenergebnisse

1. Klicken Sie in der Liste auf die letzte Kategorie, die im Zwischenergebnis berücksichtigt werden soll.
2. Klicken Sie auf **Zwischenergebnis hinzufügen....**
3. Ändern Sie optional im Dialogfeld "Zwischenergebnis definieren" den Beschriftungstext des Zwischenergebnisses.
4. Um nur ein Zwischenergebnis anzuzeigen und die Anzeige der Kategorien, durch die das Zwischenergebnis definiert wird, zu unterdrücken, wählen Sie **Zwischenergebniskategorien aus Tabelle ausblenden**.
5. Klicken Sie auf **Weiter**, um das Zwischenergebnis hinzuzufügen.

Gesamt

1. Klicken Sie auf das Kontrollkästchen **Gesamt**. Sie können außerdem den Beschriftungstext für die Gesamtsumme ändern.

Wenn die ausgewählte Variable innerhalb einer anderen Variablen verschachtelt ist, werden Gesamtsummen für jede Untertabelle eingefügt.

Anzeigeposition für Gesamtsummen und Zwischenergebnisse

Gesamtsummen und Zwischenergebnisse können über oder unter den in der jeweiligen Gesamtsumme enthaltenen Kategorien angezeigt werden.

- Wenn im Gruppenfeld "Gesamtsummen und Zwischenergebnisse erscheinen" die Option **Unter** ausgewählt ist, werden Gesamtsummen über jeder Untertabelle angezeigt und bei der Berechnung der Zwischenergebnisse werden alle Kategorien über und einschließlich der ausgewählten Kategorie (aber unter allen vorhergehenden Zwischenergebnissen) berücksichtigt.
- Wenn im Gruppenfeld "Gesamtsummen und Zwischenergebnisse erscheinen" die Option **Über** ausgewählt ist, werden Gesamtsummen unter jeder Untertabelle angezeigt und bei der Berechnung der Zwischenergebnisse werden alle Kategorien unter und einschließlich der ausgewählten Kategorie (aber über allen vorhergehenden Zwischenergebnissen) berücksichtigt.

Wichtig: Die Anzeigeposition für die Zwischenergebnisse sollte vor dem Definieren der Zwischenergebnisse ausgewählt werden. Wenn Sie die Anzeigeposition ändern, hat dies Auswirkungen auf alle Zwischenergebnisse (nicht nur auf das zurzeit ausgewählte). Außerdem werden *die in den Zwischenergebnissen enthaltenen Kategorien geändert*.

Berechnete Kategorien

Sie können Kategorien anzeigen, die aus Auswertungsstatistiken, Gesamtergebnissen, Zwischenergebnissen und/oder Konstanten berechnet wurden. Weitere Informationen finden Sie in „Berechnete Kategorien“ auf Seite 15.

Angepasste Auswertungsstatistiken für Gesamtsummen und Zwischenergebnisse

Mithilfe des Dialogfelds "Auswertungsstatistik" können Sie neben Gesamtsummen auch andere Statistiken in den Tabellenbereich für Gesamtsummen und Zwischenergebnisse aufnehmen. Weitere Informationen finden Sie in „Auswertungsstatistiken für kategoriale Variablen“ auf Seite 7.

Hinweis: Wenn Sie mehrere benutzerdefinierte Gesamtstatistiken auswählen, die sich auch im Tabellenhauptteil befinden, und die Beschriftungen für die Statistiken ausblenden, werden die Gesamtsummen in derselben Reihenfolge sortiert, in der sie sich im Tabellenkörper befinden. Da die Beschriftungen ausgeblendet sind, können Sie möglicherweise nicht erkennen, was die einzelnen Gesamtstatistiken darstellen. Im Allgemeinen empfiehlt es sich nicht, mehrere Statistiken auszuwählen und die Beschriftungen für die Statistiken auszublenden.

Gesamtsummen, Zwischenergebnisse und ausgeschlossenen Kategorien

Fälle in ausgeschlossenen Kategorien werden bei der Berechnung der Gesamtsummen nicht berücksichtigt.

Fehlende Werte, leere Kategorien und Werte ohne Wertbeschriftungen

Fehlende Werte. Hiermit wird die Anzeige von **benutzerdefiniert fehlenden** Werten bzw. von Werten gesteuert, die als fehlende Werte definiert wurden (z. B. der Code "99", der für die Schwangerschaft bei Männern "nicht zutreffend" angibt). In der Standardeinstellung werden benutzerdefiniert fehlende Werte ausgeschlossen. Wählen Sie diese Option aus (aktivieren Sie diese), um benutzerdefiniert fehlende Werte in die Tabellen aufzunehmen. Obwohl die Variable mehrere Kategorien mit fehlenden Werten enthalten kann, wird in der Tabellenvorschau im Erstellungsbereich nur eine generische Kategorie für fehlende Werte angezeigt. Alle benutzerdefiniert fehlenden Kategorien werden in die Tabelle aufgenommen. **Systemdefiniert fehlende Werte** (leere Zelle für numerische Variablen im Dateneditor) werden immer ausgeschlossen.

Leere Kategorien. Bei leeren Kategorien handelt es sich um Kategorien mit definierten Wertbeschriftungen, die aber in der betreffenden Kategorie keine Fälle für eine bestimmte Tabelle oder Untertabelle enthalten. In der Standardeinstellung werden leere Kategorien in Tabellen aufgenommen. Heben Sie die Auswahl dieser Option auf, um fehlende Kategorien aus der Tabelle auszuschließen.

Andere beim Durchsuchen der Daten gefundene Werte. In der Standardeinstellung werden Kategorie-werte in der Datendatei ohne definierte Wertbeschriftungen automatisch in die Tabelle aufgenommen. Heben Sie die Auswahl dieser Option auf, um Werte ohne definierte Wertbeschriftungen aus der Tabelle auszuschließen. Wenn Sie Kategorien mit definierten Wertbeschriftungen ausschließen, werden auch Kategorien ohne definierte Wertbeschriftungen ausgeschlossen.

Berechnete Kategorien

Neben aggregierten Ergebnissen von Auswertungsstatistiken kann eine Tabelle auch eine oder mehrere Kategorien anzeigen, die aus diesen aggregierten Ergebnissen berechnet wurden: aus konstanten Werten, Zwischenergebnissen, Gesamtsummen oder einer Kombination davon. Die Ergebnisse werden als berechnete Kategorien oder Postcomputes bezeichnet. Eine berechnete Kategorie verhält sich wie eine Kategorie in einer einzigen Variablen, wobei es folgende Ähnlichkeiten und Unterschiede gibt:

- Eine berechnete Kategorie ist wie die anderen Kategorien positioniert.
- Eine berechnete Kategorie nutzt dieselben Statistiken wie die anderen Kategorien.
- Berechnete Kategorien wirken sich nicht auf Zwischenergebnisse, Gesamtsummen oder Signifikanztests aus.
- Standardmäßig verwenden die Werte berechneter Kategorien dieselbe Formatierung für Auswertungsstatistiken wie die anderen Kategorien. Sie können das Format bei der Definition der berechneten Kategorie überschreiben.

Da berechnete Kategorien verwendet werden können, um aggregierte Ergebnisse zu summieren, können sie Zwischenergebnissen ähneln. Berechnete Kategorien haben jedoch im Vergleich zu Zwischenergebnissen folgende Vorteile:

- Berechnete Kategorien können aus den Ergebnissen anderer Zwischenergebnisse errechnet werden.
- Berechnete Kategorien können sich gegenseitig überschneiden, da sie dieselben (oder einige derselben) Kategorien nutzen.
- Berechnete Kategorien müssen nicht die Werte aller anderen Kategorien oberhalb oder unterhalb der berechneten Kategorie einschließen. Das heißt, dass berechnete Kategorien nicht abschließend sind.
- Berechnete Kategorien können Werte aus nicht benachbarten Kategorien enthalten.

Im Gegensatz zu Gesamtsummen und Zwischenergebnissen werden berechnete Kategorien anhand der aggregierten Daten berechnet, nicht anhand der ursprünglichen Daten. Deshalb ist es möglich, dass die Werte berechneter Kategorien nicht mit den Ergebnissen von Gesamtsummen und Zwischenergebnissen übereinstimmen. Da Sie außerdem die Möglichkeit haben, Quellenkategorien bei der Definition der berechneten Kategorie auszublenden, könnte es schwierig sein, die Zwischenergebnisse in der daraus resultierenden Tabelle zu interpretieren. Wenn Sie berechnete Kategorien verwenden, empfiehlt es sich, benutzerdefinierte Beschriftungen für Zwischenergebnisse anzugeben.

Definieren einer berechneten Kategorie

Berechnete Kategorien werden anhand des Dialogfelds "Kategorien und Gesamtsummen" hinzugefügt. Informationen zum Zugreifen auf dieses Dialogfeld finden Sie in „Kategorien und Gesamtsummen“ auf Seite 12.

1. Klicken Sie im Dialogfeld "Kategorien und Gesamtsummen" auf **Kategorie hinzufügen....**
2. Geben Sie unter **Beschriftung für berechnete Kategorie** eine Beschriftung für die berechnete Kategorie an. Sie können Kategorien aus der Liste "Kategorien" ziehen, um Beschriftungen für sie einzuschließen.
3. Erstellen Sie einen Ausdruck, indem Sie Kategorien und/oder Gesamtsummen und Zwischenergebnisse auswählen und mithilfe von Operatoren die berechneten Kategorien definieren. Sie können auch konstante Werte (z. B. 500) in den Ausdruck aufnehmen.
4. Um nur eine berechnete Kategorie anzuzeigen und die Anzeige der Kategorien, durch die die berechnete Kategorie definiert wird, zu unterdrücken, wählen Sie **Im Ausdruck verwendete Kategorien aus Tabelle ausblenden**.
5. Klicken Sie auf die Registerkarte **Anzeigeformate**, um das Anzeigeformat und die Anzahl an Dezimalstellen für die berechnete Kategorie zu ändern. Weitere Informationen finden Sie in „Anzeigeformate für berechnete Kategorien“.
6. Klicken Sie auf **Weiter**, um die berechnete Kategorie hinzuzufügen.

Anzeigeformate für berechnete Kategorien

Standardmäßig verwendet eine berechnete Kategorie dasselbe Anzeigeformat und dieselbe Anzahl an Dezimalstellen wie die andere Kategorie in der Variable. Sie können diese Einstellungen auf der Registerkarte "Anzeigeformate" im Dialogfeld "Berechnete Kategorie" überschreiben. Die Registerkarte "Anzeigeformate" enthält neben den Anzeigeformaten und der Anzahl an Dezimalstellen für diese Statistik auch eine Liste der aktuellen Auswertungsstatistik, die die berechnete Kategorie nutzt.

Sie haben in jeder Auswertungsstatistik folgende Möglichkeiten:

1. Wählen Sie aus der Dropdown-Liste "Format" ein Anzeigeformat für die Auswertungsstatistik aus. Eine vollständige Liste der Anzeigeformate finden Sie in „Anzeigeformate für Auswertungsstatistiken“ auf Seite 12.
2. Geben Sie in der Zelle "Dezimalstellen" die Anzahl der anzuzeigenden Dezimalstellen für die ausgewählte Auswertungsstatistik ein.

Tabellen für Variablen mit gemeinsamen Kategorien (Comperimeter-Tabellen)

Oft enthalten Umfragen viele Fragen mit übereinstimmenden möglichen Antworten. Sie können diese zusammenhängenden Variablen stapeln, damit sie in einer Tabelle dargestellt werden, und die gemeinsamen Antwortkategorien können als Spalten der Tabelle erscheinen.

So erstellen Sie eine Tabelle für mehrere Variablen mit gemeinsamen Kategorien:

Tabelle 6. Gestapelte Variablen mit gemeinsamen Antwortkategorien in Spalten

Variablen	Kategorie 1	Kategorie 2	Kategorie 3
Variable 1	12	34	56
Variable 2	56	12	34
Variable 3	34	56	12

Weitere Informationen finden Sie in Kapitel 6, „Tabellen für Variablen mit gemeinsamen Kategorien“, auf Seite 53.

Anpassen des Tabellenerstellungsprogramms

Im Gegensatz zu herkömmlichen Dialogfeldern kann die Größe des Dialogfelds für das Tabellenerstellungsprogramm wie bei Fenstern geändert werden.

1. Klicken Sie auf den oberen, unteren oder einen seitlichen Rand oder auf eine beliebige Ecke des Dialogfelds für das Tabellenerstellungsprogramm und ziehen Sie, um die Größe zu ändern.
Darüber hinaus können Sie auf der Registerkarte "Tabelle" die Größe der Variablenliste, der Liste "Kategorien" und des Erstellungsbereichs ändern.
2. Klicken Sie auf den horizontalen Balken zwischen der Variablenliste und der Liste "Kategorien" und ziehen Sie diesen, um einen längeren oder kürzeren Listenausschnitt anzuzeigen. Wenn Sie den Balken nach unten ziehen, wird der angezeigte Ausschnitt der Variablenliste länger und der Ausschnitt der Liste "Kategorien" kürzer. Wenn Sie den Balken nach oben ziehen, geschieht das Gegenteil.
3. Klicken Sie auf den vertikalen Balken zwischen der Variablenliste und der Liste "Kategorien" auf der einen und dem Erstellungsbereich auf der anderen Seite und ziehen Sie diesen, um die Listenausschnitte breiter oder schmaler zu machen. Die Größe des Erstellungsbereichs wird automatisch entsprechend dem verbleibenden Platz angepasst.

Benutzerdefinierte Tabellen: Registerkarte "Optionen"

Darstellung der Datenzelle

Legt die Anzeige von leeren Zellen und Zellen fest, für die keine Statistiken berechnet werden können.

Leere Zellen

Für Tabellenzellen, die keine Fälle enthalten (Zellenhäufigkeit 0), kann eine der folgenden drei Anzeigeeoptionen ausgewählt werden: "Null", "Leer" oder ein von Ihnen eingegebener Text. Der Text kann bis zu 255 Zeichen enthalten.

Statistiken, die nicht berechnet werden können

Text, der angezeigt wird, wenn eine Statistik nicht berechnet werden kann (z. B. der Mittelwert für eine Kategorie ohne Fälle). Der Text kann bis zu 255 Zeichen enthalten. Der Standardwert ist ein Punkt (.).

Breite der Datenspalten

Steuert die minimale und maximale Breite der Datenspalten. Diese Einstellungen haben keinen Einfluss auf die Spaltenbreite der Zeilenbeschriftung.

Einstellungen für Tabellenvorlage

Verwendet die in der Standardtabellenvorlage festgelegte Datenspaltenbreite. Sie können angepasste Standardtabellenvorlagen zur Verwendung mit neuen Tabellen erstellen sowie die Breite von Spalten für die Zeilenbeschriftung und Datenspalten mit einer Tabellenvorlage steuern.

Anpassen

Überschreibt die Einstellungen der Standardtabellenvorlage für die Breite der Datenspalten. Legt die minimale und maximale Breite der Datenspalten für die Tabelle sowie die Maßeinheit fest: Punkte, Zoll oder Zentimeter.

Fehlende Werte für metrische Variablen

Steuert die Behandlung von fehlenden Daten bei der Statistik für metrische Variablen in Tabellen mit zwei oder mehr metrischen Variablen.

Optimale Nutzung der verfügbaren Daten (variablenweiser Ausschluss)

Alle Fälle mit gültigen Werten für die einzelnen metrischen Variablen werden bei den Auswertungsstatistiken für diese metrischen Variablen berücksichtigt.

Einheitliche Fallbasis für alle metrischen Variablen (listenweiser Ausschluss)

Fälle mit fehlenden Werten für metrische Variablen in der Tabelle werden bei den Auswertungsstatistiken für alle metrischen Variablen in der Tabelle ausgeschlossen.

Effektive Basis

Sie können eine Variable, die statt Häufigkeitsgewichtungen Korrekturgewichtungen darstellt, als Gewichtungvariable für eine effektive Basis verwenden. Das Konzept der effektiven Basis oder der Gewichtung des effektiven Stichprobenumfangs basiert auf Methoden für die Analyse von Daten aus komplexen Stichproben. Die Gewichtung für eine effektive Basis ermöglicht über die Korrekturgewichtung die näherungsweise Behandlung von Statistikinferenzen in Analysen, die Ad-hoc-Anpassungen von Daten aus einfachen Zufallsstichprobendesigns beinhalten.

- Die Gewichtung für eine effektive Basis wirkt sich auf die gewichteten Werte in den Auswertungsstatistiken und auf die Spaltenmittel- und Spaltenanteilsignifikanztests aus.
- Wenn die Gewichtung für ein Dataset aktiviert ist, wird die Gewichtungvariable für Datasets ignoriert und die Ergebnisse werden durch die Gewichtungvariable für eine effektive Basis gewichtet.
- Die Gewichtungvariable für eine effektive Basis muss numerisch sein.
- Fälle mit negativen Gewichtungswerten, dem Gewichtungswert 0 oder fehlenden Gewichtungswerten werden von allen Ergebnissen ausgeschlossen.

Doppeltantworten für Sets aus kategorialen Variablen zählen

Bei einer Doppeltantwort wurde für zwei oder mehr Variablen in einem Set von kategorialen Variablen dieselbe Antwort gegeben. Standardmäßig werden Doppeltantworten nicht gezählt.

Kleine Werte für Häufigkeiten ausblenden

Sie können festlegen, dass Häufigkeiten ausgeblendet werden, wenn sie einen bestimmten ganzzahligen Wert unterschreiten.

- Ausgeblendete Werte werden als $<N$ angezeigt. Dabei ist N die angegebene Ganzzahl.
- Die angegebene Ganzzahl muss größer oder gleich 2 sein.
- Wenn die Gewichtung für das Dataset aktiviert ist oder eine Gewichtungvariable für eine effektive Basis angegeben ist, wird der gewichtete Wert verwendet.

Gewichtungen und Rundung

- Wenn Sie Fälle über **Daten > Fälle gewichten** gewichten, werden nicht ganzzahlige Gewichtungen auf Zellen- oder Kategorieebene für Signifikanztests, Konfidenzintervalle und Standardfehler gerundet.
- Wenn Sie **Gewichtungvariable für effektive Basis verwenden** auswählen, werden nicht ganzzahlige Gewichtungswerte in der ausgewählten GewichtungsvARIABLEN nicht gerundet.

- Wenn beide Optionen angegeben wurden, wird die GewichtungsvARIABLE für die effektive Basis verwendet und nicht ganzzahlige Gewichtungen werden nicht gerundet.

Benutzerdefinierte Tabellen: Registerkarte "Titel"

Auf der Registerkarte "Titel" wird die Anzeige von Titeln, Titelzeilen und Eckenbeschriftungen gesteuert.

Titel. Text, der über der Tabelle angezeigt wird.

Titelzeile. Text, der unter der Tabelle und, sofern vorhanden, über den Fußnoten angezeigt wird.

Ecke. Text, der links oben in der Ecke der Tabelle angezeigt wird. Der Eckentext wird nur dann angezeigt, wenn die Tabelle Zeilenvariablen enthält und die Eigenschaft "Beschriftung für Zeilendimension" für die Pivot-Tabelle auf **Verschachtelt** gesetzt ist. In der Standardtabellenvorlage ist dies *nicht* der Fall.

Die folgenden automatisch erstellten Werte können in den Tabellentitel, die Titelzeile oder den Eckenbeschriftung aufgenommen werden:

Datum. Aktuelles Datum mit Jahr, Monat und Tag, angezeigt in dem Format, das in den regionalen Einstellungen bzw. Ländereinstellungen von Windows festgelegt ist.

Uhrzeit. Aktuelle Uhrzeit mit Stunde, Minute und Sekunde, angezeigt in dem Format, das in den regionalen Einstellungen bzw. Ländereinstellungen von Windows festgelegt ist.

Tabellenausdruck. In der Tabelle verwendete Variablen und deren Verwendung in der Tabelle. Wenn eine Variable über eine definierte Variablenbeschriftung verfügt, wird die Beschriftung angezeigt. Die folgenden Symbole zeigen in der erstellten Tabelle an, wie Variablen in dieser verwendet werden.

- + zeigt gestapelte Variablen an.
- > zeigt Verschachtelung an.
- Nach zeigt Kreuztabellen oder Schichten an.

Benutzerdefinierte Tabellen: Registerkarte "Teststatistiken"

Auf der Registerkarte **Teststatistiken** werden Signifikanztests für benutzerdefinierte Tabellen bereitgestellt.

Diese Tests sind weder für Tabellen, bei denen Kategoriebeschriftungen aus den Standardtabellendimensionen verschoben wurden, noch für berechnete Kategorien verfügbar.

Spaltenmittel- und Spaltenanteiletests

Spaltenmitteltests stehen für metrische Variablen zur Verfügung. Spaltenanteiletests stehen für kategoriale Variablen zur Verfügung.

Spaltenmittel vergleichen

Paarweiser Test auf Gleichheit der Spaltenmittel. Die Tabelle muss eine kategoriale Variable in den Spalten enthalten und eine metrische Variable als innerste Ebene der Zeilen. Die Tabelle muss den Mittelwert als Auswertungsstatistik beinhalten.

Die Varianz kann für reguläre kategoriale Variablen auf der Grundlage aller Kategorien oder auf der Grundlage der verglichenen Kategorien geschätzt werden. Für Mehrfachantwortvariablen wird die Varianz für den Spaltenmitteltest immer nur auf der Grundlage der verglichenen Kategorien geschätzt.

Spaltenanteile vergleichen

Paarweiser Test auf Gleichheit der Spaltenanteile. Die Tabelle muss mindestens eine kategoriale Variable sowohl in den Spalten als auch in den Zeilen enthalten. Die Tabelle muss Häufigkeiten oder Spaltenprozentage beinhalten.

Signifikante Unterschiede identifizieren

Für Spaltenmittel- und Spaltenanteiletests können Sie signifikante Ergebnisse in einer separaten Tabelle oder in der Haupttabelle anzeigen.

In einer eigenen Tabelle

Ergebnisse von Signifikanztests werden in einer separaten Tabelle angezeigt. Falls zwei Werte deutlich voneinander abweichen, wird in der dem höheren Wert zugehörigen Zelle ein Schlüssel angezeigt, der auf die Spalte mit dem kleineren Wert verweist.

Signifikanzwerte anzeigen

Die Signifikanzwerte werden nach jedem Schlüsselwert in der Zelle in Klammern angezeigt. Diese Option ist nur verfügbar, wenn signifikante Ergebnisse in einer separaten Tabelle angezeigt werden.

In der Haupttabelle

Ergebnisse von Signifikanztests werden in der Haupttabelle angezeigt. Jede Spaltenkategorie in der Tabelle wird durch einen alphabetischen Schlüssel gekennzeichnet. Für jedes signifikante Paar wird der Schlüssel der Kategorie mit dem kleineren Spaltenmittel oder Spaltenanteil in der Kategorie mit dem größeren Spaltenmittel oder Spaltenanteil angezeigt.

- Wenn Sie mit der Maus über einen Schlüssel in der Spaltenbeschriftungszelle in einer Pivot-Tabelle fahren, werden alle Zellen in der Tabelle mit diesem Signifikanzschlüssel hervorgehoben. Für eine Tabelle mit mehreren Variablen in der Spaltendimension werden nur die Zellen in dieser Untertabelle hervorgehoben.
- Wenn Sie alle Zellen in einer Tabelle (bzw. Untertabelle) auswählen wollen, die denselben Signifikanzschlüssel aufweisen, klicken Sie mit der rechten Maustaste auf die Spaltenbeschriftungszelle und wählen **Auswählen > Alle Zellen mit diesem Signifikanzschlüssel auswählen** aus.

Tiefgestellte Zeichen im APA-Stil

Signifikante Unterschiede werden durch eine Formatierung im APA-Stil angegeben, bei der tiefgestellte Buchstaben verwendet werden. Falls zwei Werte deutlich voneinander abweichen, weisen diese Werte unterschiedliche tiefgestellte Buchstaben auf. Diese tiefgestellten Buchstaben sind keine Fußnoten. Ist diese Option aktiviert, wird der in der aktuellen Tabellenvorlage definierte Fußnotenstil überschrieben und tatsächliche Fußnoten werden als hochgestellte Zahlen angezeigt. Wenn Sie alle Zellen in derselben Zeile auswählen wollen, die denselben Signifikanzschlüssel aufweisen, klicken Sie mit der rechten Maustaste auf eine Zelle, die einen Signifikanzschlüssel aufweist, und wählen **Zellen mit ähnlicher Signifikanz auswählen** aus.

Signifikanzniveaus

Das Signifikanzniveau für Spaltenmittel- und Spaltenanteiletests.

- Der Wert muss größer als 0 und kleiner als 1 sein.
- Wenn Sie zwei Signifikanzniveaus angeben, werden Großbuchstaben zur Identifizierung der Signifikanzwerte kleiner-gleich dem niedrigeren Niveau verwendet. Kleinbuchstaben werden zur Identifizierung der Signifikanzwerte kleiner-gleich dem höheren Niveau verwendet.
- Wenn Sie **Tiefgestellte Zeichen im APA-Stil** auswählen, wird der zweite Wert ignoriert.

p-Werte für Mehrfachvergleiche anpassen

Bei der **Bonferroni**-Korrektur werden Anpassungen für die Familywise Error Rate (FWER) vorgenommen. Die **Benjamini-Hochberg**-Methode ist eine Anpassung für die False Discovery Rate (FDR). Diese Methode ist weniger konservativ als die Bonferroni-Korrektur.

Tests auf Unabhängigkeit (Chi-Quadrat)

Chi-Quadrat-Test auf Unabhängigkeit für Tabellen, in denen sowohl in den Zeilen als auch in den Spalten mindestens eine Kategorievariable vorhanden ist.

Zwischensummen anstelle von Kategorien für Zwischensummen verwenden

In Signifikanztests werden statt den Kategorien für Zwischenergebnisse die Zwischenergebnisse selbst verwendet. Andernfalls werden nur die für Zwischenergebnisse verwendeten ausgeblendeten Kategorien in den Tests durch die Zwischenergebnisse selbst ersetzt.

Mehrfachantwortvariablen in Tests einschließen

Kategorien von Mehrfachantwortsets werden in Signifikanztests eingeschlossen. Andernfalls werden Mehrfachantwortsets nicht in Signifikanztests eingeschlossen.

Kapitel 2. Einfache Tabellen für kategoriale Variablen

Einfache Tabellen für kategoriale Variablen

Die meisten Tabellen, die Sie erstellen, enthalten wahrscheinlich mindestens eine **kategoriale Variable**. Eine kategoriale Variable ist eine Variable mit einer bestimmten Anzahl von eindeutigen Werten bzw. Kategorien (beispielsweise Geschlecht oder Religion). Kategoriale Variablen können **nominal** oder **ordinal** sein.

- *Nominal*. Eine Variable kann als nominal behandelt werden, wenn ihre Werte Kategorien darstellen, die sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Firmenabteilung, in der eine Person arbeitet. Beispiele für nominale Variablen sind Region, Postleitzahl oder Religionszugehörigkeit.
- *Ordinal*. Eine Variable kann als ordinal behandelt werden, wenn ihre Werte für Kategorien stehen, die eine natürliche Reihenfolge aufweisen (z. B. Grad der Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden). Ordinale Variablen treten beispielsweise bei Einstellungsmessungen (Zufriedenheit oder Vertrauen) und bei Präferenzbeurteilungen auf.

Der Variablentyp ist durch ein Symbol neben der jeweiligen Variablen in der Variablenliste gekennzeichnet.

Tabelle 7. Messniveausymbole

	Numerisch	Zeichenfolge	Datum	Zeit
Metrisch (stetig)		entfällt		
Ordinal				
Nominal				

Die Funktion zum Erstellen von benutzerdefinierten Tabellen wurde für kategoriale Variablen mit definierten **Wertbeschriftungen** optimiert. Weitere Informationen finden Sie in „Erstellen von Tabellen“ auf Seite 1.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Einzelne kategoriale Variable

Eine Tabelle mit einer einzelnen kategorialen Variablen ist wahrscheinlich eine der einfachsten Tabellen, die Sie erstellen können. Oft werden Sie jedoch keine komplexere Tabelle benötigen.

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

- Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.

Im Erstellungsbereich wird eine Vorschau der Tabelle angezeigt. In der Vorschau werden nicht die tatsächlichen Datenwerte angezeigt, sondern Platzhalter anstelle der eigentlichen Daten.

- Klicken Sie auf **OK**, um die Tabelle zu erstellen.

Die Tabelle wird im Fenster des Viewers angezeigt.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Abbildung 2. Eine kategoriale Variable in Zeilen

In dieser einfachen Tabelle ist die Spaltenüberschrift *Anzahl* eigentlich nicht notwendig und die Tabelle kann auch ohne diese Spaltenüberschrift erstellt werden.

- Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
- Aktivieren Sie im Gruppenfeld "Auswertungsstatistik" neben der Dropdown-Liste "Position" die Option **Ausblenden**.
- Klicken Sie auf **OK**, um die Tabelle zu erstellen.

Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Abbildung 3. Eine kategoriale Variable ohne Spaltenbeschriftung für die Auswertungsstatistik

Prozentsätze

Zusätzlich zu Häufigkeiten können Sie auch Prozentsätze anzeigen lassen. Bei einer einfachen Tabelle mit einer einzelnen kategorialen Variablen, die in Zeilen angezeigt wird, werden wahrscheinlich die Spaltenprozentsätze interessant sein. Wenn die Variable hingegen in Spalten angezeigt wird, interessieren Sie wahrscheinlich eher die Zeilenprozentsätze.

- Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
- Heben Sie im Gruppenfeld "Auswertungsstatistik" neben der Dropdown-Liste "Position" die Auswahl der Option **Ausblenden** auf. Da diese Tabelle zwei Spalten enthält, sollen die Spaltenbeschriftungen angezeigt werden, sodass Sie wissen, welche Größen in den Spalten dargestellt werden.
- Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus.
- Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
- Löschen Sie in der Liste "Anzeigen" in der Zelle mit der Beschriftung die Standardbeschriftung und geben Sie Prozent ein.
- Klicken Sie auf **Der Auswahl zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%

Abbildung 4. Häufigkeiten und Spaltenprozentsätze

Gesamtsummen

Gesamtsummen werden nicht automatisch in benutzerdefinierte Tabellen aufgenommen. Gesamtsummen können einer Tabelle jedoch problemlos hinzugefügt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken.
4. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
Total		2828	100.0%

Abbildung 5. Häufigkeiten, Spaltenprozentsätze und Gesamtsummen

Weitere Informationen finden Sie in Kapitel 4, „Gesamtsummen und Zwischenergebnisse für kategoriale Variablen“, auf Seite 39.

Kreuztabelle

Eine Kreuztabelle stellt ein grundlegendes Verfahren dar, mit dem die Beziehung zwischen zwei kategorialen Variablen untersucht werden kann. Wenn Sie beispielsweise *Alterskategorie* als Zeilenvariable und *Geschlecht* als Spaltenvariable verwenden, können Sie eine zweidimensionale Kreuztabelle erstellen, mit der die Anzahl an Männern und Frauen in jeder Alterskategorie dargestellt wird.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
4. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste in den Spaltenbereich im Erstellungsbereich. (Möglicherweise müssen Sie in der Variablenliste blättern, um diese Variable zu finden.)
5. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Gender	
		Male	Female
		Count	Count
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

Abbildung 6. Kreuztabelle von Alterskategorie und Geschlecht

Prozentsätze in Kreuztabellen

In einer zweidimensionalen Kreuztabelle sind möglicherweise sowohl die Zeilen- als auch die Spaltenprozentsätze nützliche Informationen.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Geschlecht*.
Möglicherweise fällt Ihnen hierbei auf, dass die Option **Auswertungsstatistik** im Popup-Menü inaktiviert ist. Der Grund hierfür ist, dass Sie Auswertungsstatistiken nur für die innerste Variable in der Quellendimension der Statistik auswählen können. Die Standardquellendimension der Statistik (Zeile oder Spalte) für kategoriale Variablen beruht auf der Reihenfolge, in der Sie die Variablen in den Erstellungsbereich ziehen. In diesem Beispiel wurde *Alterskategorie* zuerst in die Zeilendimension gezogen. Da dort keine weiteren Variablen vorliegen, ist *Alterskategorie* die Quellenvariable der Statistik. Sie können die Quellendimension der Statistik ändern. In diesem Beispiel ist dies jedoch nicht nötig. Weitere Informationen finden Sie in „Auswertungsstatistik“ auf Seite 6.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
4. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
5. Wählen Sie in der Liste "Statistik" die Option **Anzahl als Zeilen%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
6. Klicken Sie auf **Der Auswahl zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65 or older	178	14.5%	37.2%	301	18.8%	62.8%

Abbildung 7. Kreuztabelle mit Zeilen- und Spaltenprozentsätzen

Steuern des Anzeigeformats

Sie können das Anzeigeformat einschließlich der Anzahl der in Auswertungsstatistiken angezeigten Dezimalstellen steuern. Prozentsätze werden in der Standardeinstellung beispielsweise mit einer Dezimalstelle und einem Prozentzeichen angezeigt. Möglicherweise möchten Sie die Zellenwerte jedoch mit zwei Dezimalstellen, aber ohne Prozentzeichen darstellen.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
3. Wählen Sie für die beiden ausgewählten Auswertungsstatistiken mit Prozentsätzen (**Anzahl als Spalten%** und **Anzahl als Zeilen%**) in der Dropdown-Liste "Format" die Option **nnnn.n** aus und geben Sie für beide in der Zelle "Dezimalstellen" den Wert 2 ein.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.79	44.63	134	8.38	55.37
	25 to 34	276	22.48	44.02	351	21.94	55.98
	35 to 44	309	25.16	45.51	370	23.13	54.49
	45 to 54	221	18.00	45.95	260	16.25	54.05
	55 to 64	136	11.07	42.50	184	11.50	57.50
	65 or older	178	14.50	37.16	301	18.81	62.84

Abbildung 8. Formatierte Zellenanzeige für Zeilen- und Spaltenprozentsätze

Randsummen

In Kreuztabellen sollen häufig **Randsummen** (Gesamtsummen für jede Zeile und Spalte) angezeigt werden. Da diese in der Standardeinstellung nicht in benutzerdefinierten Tabellen enthalten sind, müssen sie den Tabellen ausdrücklich hinzugefügt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
4. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste in den Spaltenbereich im Erstellungsbereich. (Möglicherweise müssen Sie in der Variablenliste blättern, um diese Variable zu finden.)
5. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
6. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken, und klicken Sie anschließend auf **Zuweisen**.
7. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Geschlecht* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
8. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken, und klicken Sie anschließend auf **Zuweisen**.
9. Aktivieren Sie im Gruppenfeld "Auswertungsstatistik" neben der Dropdown-Liste "Position" die Option **Ausblenden**. (Da nur Häufigkeiten angezeigt werden, muss die in den Datenzellen der Tabelle angezeigte "Statistik" nicht angegeben werden.)
10. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Gender		
		Male	Female	Total
Age category	Less than 25	108	134	242
	25 to 34	276	351	627
	35 to 44	309	370	679
	45 to 54	221	260	481
	55 to 64	136	184	320
	65 or older	178	301	479
	Total	1228	1600	2828

Abbildung 9. Kreuztabelle mit Randsummen

Sortieren und Ausschließen von Kategorien

In der Standardeinstellung werden die Kategorien in aufsteigender Reihenfolge der Datenwerte, die durch die Wertbeschriftungen der Kategorien dargestellt werden, sortiert. Obwohl für Alterskategorien beispielsweise die Wertbeschriftungen *Unter 25*, *25 bis 34*, *35 bis 44* usw. angezeigt werden, lauten die tatsächlich zugrunde liegenden Datenwerte 1, 2, 3 usw. Die Standardanzweigerihenfolge der Kategorien wird durch diese zugrunde liegenden Datenwerte festgelegt.

Die Reihenfolge der Kategorien kann leicht geändert werden und Kategorien, die in der Tabelle nicht angezeigt werden sollen, können ausgeschlossen werden.

Sortieren von Kategorien

Sie können Kategorien manuell neu anordnen oder Kategorien in aufsteigender oder absteigender Reihenfolge nach folgenden Elementen sortieren:

- Datenwerte.
 - Wertbeschriftungen.
 - Zellenhäufigkeit.
 - Auswertungsstatistik. Welche Auswertungsstatistik für die Sortierung verfügbar ist, hängt davon ab, welche Auswertungsstatistik Sie für die Anzeige in der Tabelle ausgewählt haben.
1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
 2. Wenn die Variable *Alterskategorie* nicht bereits im Erstellungsbereich im Bereich für die Zeilen angezeigt wird, ziehen Sie sie dorthin und legen sie dort ab.
 3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
Sowohl die Datenwerte als auch die damit verbundenen Wertbeschriftungen werden in der aktuellen Sortierreihenfolge angezeigt. In diesem Fall handelt es hierbei noch um die aufsteigende Reihenfolge der Datenwerte.
 4. Wählen Sie im Gruppenfeld "Kategorien sortieren" in der Dropdown-Liste "Reihenfolge" die Option **Absteigend** aus.
Die Sortierreihenfolge wurde nun umgekehrt.
 5. Wählen Sie in der Dropdown-Liste "Nach" die Option **Beschriftungen** aus.
Die Kategorien werden nun in absteigender alphabetischer Reihenfolge der Wertbeschriftungen sortiert.
Beachten Sie, dass sich die Kategorie mit der Beschriftung *Unter 25* am Anfang der Liste befindet. Bei der alphabetischen Reihenfolge werden Buchstaben nach Ziffern angezeigt. Da dies die einzige Beschriftung ist, die mit einem Buchstaben anfängt und die Liste in absteigender (umgekehrter) Reihenfolge sortiert wird, wird diese Kategorie am Anfang der Liste angezeigt.

Wenn eine bestimmte Kategorie an anderer Stelle in der Liste angezeigt werden soll, können Sie sie leicht dorthin verschieben.

6. Klicken Sie in der Liste "Beschriftung" auf die Kategorie mit der Beschriftung *Unter 25*.
7. Klicken Sie auf den Abwärtspfeil rechts neben der Liste. Die Kategorie wird nun in der Liste um eine Zeile nach unten verschoben.
8. Klicken Sie so oft auf den Abwärtspfeil, bis sich die Kategorie am Ende der Liste befindet.

Ausschließen von Kategorien

Kategorien, die nicht in der Tabelle angezeigt werden sollen, können ausgeschlossen werden.

1. Klicken Sie in der Liste "Beschriftung" auf die Kategorie mit der Beschriftung *Unter 25*.
2. Klicken Sie auf die Pfeilschaltfläche links neben der Liste "Ausschließen".
3. Klicken Sie in der Liste "Beschriftung" auf die Kategorie mit der Beschriftung *65 oder älter*.
4. Klicken Sie erneut auf die Pfeilschaltfläche links neben der Liste "Ausschließen".

Daraufhin werden die beiden Kategorien aus der Liste "Anzeigen" in die Liste "Ausschließen" verschoben. Wenn Sie dies rückgängig machen möchten, können Sie sie leicht wieder in die Liste "Anzeigen" aufnehmen.

5. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Gender		
		Male	Female	Total
Age category	55 to 64	136	184	320
	45 to 54	221	260	481
	35 to 44	309	370	679
	25 to 34	276	351	627
	Total	942	1165	2107

Abbildung 10. Nach absteigender Wertbeschriftung sortierte Tabelle mit einigen ausgeschlossenen Kategorien

Beachten Sie, dass die Gesamtsummen niedriger als vor dem Ausschluss der beiden Kategorien sind. Dies liegt daran, dass Gesamtsummen auf den in der Tabelle enthaltenen Kategorien beruhen. Ausgeschlossene Kategorien werden auch aus der Berechnung der Gesamtsummen ausgeschlossen. Weitere Informationen finden Sie in Kapitel 4, „Gesamtsummen und Zwischenergebnisse für kategoriale Variablen“, auf Seite 39.

Kapitel 3. Stapeln, Verschachteln und Schichten von kategorialen Variablen

Stapeln, Verschachteln und Schichten sind Methoden, um in derselben Tabelle mehrere Variablen gleichzeitig anzuzeigen. In diesem Kapitel wird die Verwendung dieser Verfahren bei kategorialen Variablen behandelt. Die Verfahren können jedoch auch bei metrischen Variablen verwendet werden.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Stapeln von kategorialen Variablen

Beim Stapeln werden separate Tabellen in dieselbe Anzeige eingefügt. So können beispielsweise Informationen über das *Geschlecht* und die *Alterskategorie* in unterschiedlichen Abschnitten derselben Tabelle angezeigt werden.

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

2. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste des Tabellenerstellungsprogramms in den Zeilenbereich im Erstellungsbereich.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste und legen Sie sie im Zeilenbereich unter *Geschlecht* ab.

Die beiden Variablen sind nun in der Zeilendimension gestapelt.

4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count
Gender	Male	1232
	Female	1600
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Abbildung 11. Tabelle mit kategorialen Variablen, die in den Zeilen gestapelt sind

Auf dieselbe Weise können Variablen in den Spalten gestapelt werden.

Stapeln bei Kreuztabellen

Eine gestapelte Tabelle kann andere Variablen in anderen Dimensionen enthalten. So können beispielsweise zwei in den Zeilen gestapelte Variablen mit einer dritten Variable, die in der Spaltendimension dargestellt ist, zu einer Kreuztabelle kombiniert werden.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabelle", Befehl "Benutzerdefinierte Tabellen").

2. Falls *Alterskategorie* und *Geschlecht* nicht bereits in den Zeilen gestapelt sind, folgen Sie den vorangegangenen Anweisungen zum Stapeln von Variablen.
3. Ziehen Sie die Variable *Nachrichten aus dem Internet* aus der Variablenliste in den Erstellungsbereich und legen Sie im Spaltenbereich ab.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Get news from internet	
		No	Yes
		Count	Count
Gender	Male	873	359
	Female	1092	508
Age category	Less than 25	146	96
	25 to 34	368	259
	35 to 44	435	244
	45 to 54	346	135
	55 to 64	252	68
	65 or older	416	63

Abbildung 12. Zwei gestapelte Zeilenvariablen wurden mit einer Spaltenvariablen zu einer Kreuztabelle kombiniert.

Hinweis: Es sind mehrere Variablen vorhanden, deren Beschriftungen mit dem Wort *Nachrichten* beginnen, sodass es schwierig sein könnte, diese in der Variablenliste zu unterscheiden (da die Beschriftungen für eine vollständige Anzeige in der Variablenliste möglicherweise zu lang sind). Es gibt zwei Möglichkeiten, die Beschriftungen der Variablen vollständig anzuzeigen:

- Zeigen Sie mit dem Mauszeiger auf eine Variable in der Liste, um die vollständige Beschriftung als QuickInfo einzublenden.
- Klicken Sie auf den vertikalen Balken zwischen der Variablenliste und der Liste "Kategorien" auf der einen und dem Erstellungsbereich auf der anderen Seite und ziehen Sie diesen, um die Listenausschnitte breiter zu machen.

Verschachteln von kategorialen Variablen

Ähnlich wie bei Kreuztabellen kann mit Verschachtelung die Beziehung zwischen zwei kategorialen Variablen aufgezeigt werden. Bei der Verschachtelung wird eine der Variablen jedoch in derselben Dimension innerhalb der anderen Variablen verschachtelt. So können Sie beispielsweise *Geschlecht* in *Alterskategorie* in der Zeilendimension verschachteln, um die Anzahl der Männer und Frauen in jeder Alterskategorie darzustellen.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
4. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste und legen Sie sie im Zeilenbereich rechts neben *Alterskategorie* ab.

In der Vorschau im Erstellungsbereich wird nun eine verschachtelte Tabelle angezeigt, die eine einzelne Spalte mit Häufigkeiten enthält. Die Zellen der Spalte enthalten die Anzahl der Frauen und Männer in der jeweiligen Alterskategorie.

Es wird Ihnen auffallen, dass die Variablenbeschriftung *Geschlecht* bei jeder Alterskategorie angezeigt wird. Solche Wiederholungen können reduziert werden, wenn Sie die Variable mit der geringsten Anzahl von Kategorien auf der äußersten Verschachtelungsebene anordnen.

5. Klicken Sie im Erstellungsbereich auf die Variablenbeschriftung *Geschlecht*.
6. Ziehen Sie die Variable und legen Sie sie so weit links wie möglich im Zeilenbereich ab.

Die Beschriftung *Geschlecht* wird nun nicht mehr sechsmal wiederholt, *Alterskategorie* dafür zweimal. Diese Tabelle ist übersichtlicher und stellt im Wesentlichen dieselben Ergebnisse dar.

7. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

				Count
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age category	Less than 25	134
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

Abbildung 13. Tabelle der Alterskategorie verschachtelt im Geschlecht.

Hinweis: Bei benutzerdefinierten Tabellen wird die Verarbeitung geschichteter aufgeteilter Dateien nicht berücksichtigt. Um dasselbe Ergebnis wie bei geschichteten aufgeteilten Dateien zu erzielen, müssen Sie die Dateiteilungsvariablen in den äußersten Verschachtelungsebenen der Datei platzieren.

Unterdrücken von Variablenbeschriftungen

Eine andere Möglichkeit zum Vermeiden von überflüssigen Variablenbeschriftungen in verschachtelten Tabellen besteht darin, die Anzeige von Variablennamen oder -beschriftungen einfach zu unterdrücken. Da die Wertbeschriftungen von *Geschlecht* und *Alterskategorie* selbsterklärend sein dürften, können die Beschriftungen für beide Variablen entfernt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und heben Sie im Pop-up-Menü die Auswahl von **Variablenbeschriftung anzeigen** auf.
3. Wiederholen Sie diesen Vorgang für *Geschlecht*.

Die Variablenbeschriftungen werden weiterhin in der Tabellenvorschau angezeigt, sind aber nicht mehr in der Tabelle enthalten.

4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Abbildung 14. Verschachtelte Tabelle ohne Variablenbeschriftungen

Wenn die Variablenbeschriftungen in die Tabelle aufgenommen werden sollen, ohne dass diese mehrfach wiederholt im Tabellenkorpus angezeigt werden, können Sie sie im Tabellentitel oder in der Eckenbeschriftung unterbringen.

5. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").

6. Klicken Sie auf die Registerkarte **Titel**.
7. Klicken Sie im Textfeld "Titel" auf eine beliebige Stelle.
8. Klicken Sie auf **Tabellenausdruck**. Im Textfeld "Titel" wird der Text *&[Tabellenausdruck]* angezeigt. Dies ergibt einen Tabellentitel, der die Variablenbeschriftungen für die in der Tabelle verwendeten Variablen enthält.
9. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

Gender > Age category

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Abbildung 15. Variablenbeschriftungen im Tabellentitel

Das Größer-als-Zeichen (>) im Titel zeigt an, dass *Alterskategorie* in *Geschlecht* verschachtelt ist.

Verschachtelte Kreuztabellen

Eine verschachtelte Tabelle kann andere Variablen in anderen Dimensionen enthalten. Sie können beispielsweise *Alterskategorie* in *Geschlecht* in Zeilen verschachteln und diese verschachtelten Zeilen mit einer dritten Variablen in der Spaltendimension zu einer Kreuztabelle kombinieren.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Falls *Alterskategorie* noch nicht in den Zeilen in *Geschlecht* verschachtelt ist, befolgen Sie die Anweisungen im vorherigen Beispiel zum Verschachteln von Variablen.
3. Ziehen Sie die Variable *Nachrichten aus dem Internet* aus der Variablenliste in den Erstellungsbereich und legen Sie im Spaltenbereich ab.

Sie werden feststellen, dass die Tabelle zu groß ist, um vollständig im Erstellungsbereich angezeigt zu werden. Sie können im Erstellungsbereich nach oben/unten bzw. rechts/links blättern, um andere Bereiche der Tabellenvorschau einzublenden, oder folgende Aktionen durchführen:

- Klicken Sie im Tabellenerstellungsprogramm auf **Kompakt**, um eine Kompaktansicht anzuzeigen. Hierbei werden nur die Variablenbeschriftungen ohne Informationen zu den in der Tabelle enthaltenen Kategorien oder Auswertungsstatistiken angezeigt.
 - Vergrößern Sie das Dialogfeld des Tabellenerstellungsprogramms, indem Sie auf eine beliebige Seite oder Ecke des Dialogfelds des Tabellenerstellungsprogramms klicken und diese ziehen.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

				Get news from internet	
				No	Yes
				Count	Count
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
			35 to 44	217	92
			45 to 54	169	52
			55 to 64	112	24
			65 or older	155	23
	Female	Age category	Less than 25	87	47
			25 to 34	209	142
			35 to 44	218	152
			45 to 54	177	83
			55 to 64	140	44
			65 or older	261	40

Abbildung 16. Verschachtelte Kreuztabelle

Vertauschen von Zeilen und Spalten

Wie gehen Sie vor, wenn Sie viel Zeit aufgewendet haben, um genau die gewünschte Tabelle zu erstellen, dann aber feststellen, dass die Orientierung der Tabelle geändert werden müsste, indem die Zeilenvariablen in die Spalten und die Spaltenvariablen in die Zeilen verschoben werden? Angenommen, Sie haben eine verschachtelte Kreuztabelle erstellt, bei der *Alterskategorie* und *Geschlecht* in den Zeilen verschachtelt sind. Nun möchten Sie diese beiden demografischen Variablen jedoch lieber in den Spalten verschachteln.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf eine beliebige Stelle und wählen Sie im Popup-Menü die Option **Zeilen- und Spaltenvariablen vertauschen** aus.
Die Zeilen- und Spaltenvariablen sind nun vertauscht.
Vor dem Erstellen der Tabelle sollten noch einige Änderungen vorgenommen werden, um die Anzeige etwas übersichtlicher zu gestalten.
3. Wählen Sie **Verbergen** aus, um die Spaltenbeschriftungen für Auswertungsstatistiken zu unterdrücken.
4. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Geschlecht* und heben Sie die Auswahl von **Variablenbeschriftung anzeigen** auf.
5. Klicken Sie jetzt auf **OK**, um die Tabelle zu erstellen.

		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

Abbildung 17. Kreuztabelle mit in Spalten verschachtelten demografischen Variablen

Schichten

Mit Schichten können Tabellen um eine Tiefendimension erweitert werden, um dreidimensionale "Würfel" zu erstellen. Schichten ähneln stark der Verschachtelung oder Stapelung. Der wichtigste Unterschied besteht darin, dass stets nur eine Schichtenkategorie sichtbar ist. Wenn Sie beispielsweise *Alterskategorie* als Zeilenvariable und *Geschlecht* als Schichtvariable verwenden, erhalten Sie eine Tabelle, in der die Informationen für Männer und Frauen in unterschiedlichen Schichten dargestellt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").

2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
4. Klicken Sie im oberen Bereich des Dialogfelds des Tabellenerstellungsprogramm auf **Schichten**, um die Liste "Schichten" anzuzeigen.
5. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste und legen Sie diese in der Liste "Schichten" ab.

Sie werden an dieser Stelle feststellen, dass sich das Hinzufügen einer Schichtvariable in der Vorschau im Erstellungsbereich optisch nicht bemerkbar macht. Schichtvariablen haben nur dann Auswirkungen auf die Vorschau im Erstellungsbereich, wenn die Schichtvariable die Quellenvariable für die Statistik ist und Sie die Auswertungsstatistik ändern.

6. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count
Age category	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178

Abbildung 18. Einfache Tabelle mit Schichten

Auf den ersten Blick unterscheidet sich diese Tabelle nicht von einer einfachen Tabelle mit einer einzelnen kategorialen Variablen. Der einzige Unterschied besteht im Vorhandensein der Beschriftung *Geschlecht männlich* am oberen Rand der Tabelle.

7. Doppelklicken Sie im Fenster des Viewers auf die Tabelle, um diese zu aktivieren.
8. Nun wird klar, dass es sich bei der Beschriftung *Geschlecht männlich* eigentlich um einen Eintrag in einer Dropdown-Liste handelt.
9. Klicken Sie auf den Abwärtspfeil der Dropdown-Liste, um die vollständige Liste der Schichten anzuzeigen.

Bei dieser Tabelle enthält die Liste nur einen weiteren Eintrag.

10. Wählen Sie in der Dropdown-Liste den Eintrag *Geschlecht weiblich* aus.

		Count
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Abbildung 19. Einfache Tabelle mit Schichten; angezeigt wird die andere Schicht

Zwei gestapelte kategoriale Schichtvariablen

Wenn die Schichten mehrere kategoriale Variablen enthalten, können Sie die Schichtvariablen stapeln oder verschachteln. In der Standardeinstellung werden Schichtvariablen gestapelt. (*Hinweis:* Bei metrischen Schichtvariablen ist nur Stapeln möglich.)

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").

2. Wenn sich *Alterskategorie* nicht bereits in den Zeilen und *Geschlecht* nicht bereits in den Schichten befindet, befolgen Sie die Anweisungen im vorherigen Beispiel zum Erstellen einer Tabelle mit Schichten.
3. Ziehen Sie die Variable *Höchster Schulabschluss* aus der Variablenliste und legen Sie sie in der Liste "Schichten" unter *Geschlecht* ab.
Die beiden Optionsfelder im Gruppenfeld "Schichtausgabe" unter der Liste "Schichten" sind jetzt freigegeben. In der Standardeinstellung ist **Jede Kategorie als Schicht zeigen** aktiviert. Dies entspricht der Stapelung.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.
5. Doppelklicken Sie im Fenster des Viewers auf die Tabelle, um diese zu aktivieren.
6. Klicken Sie auf den Abwärtspfeil der Dropdown-Liste, um die vollständige Liste der Schichten anzuzeigen.

Die Tabelle enthält sieben Schichten: zwei Schichten für die beiden Kategorien von *Geschlecht* und fünf Schichten für die fünf Kategorien von *Höchster Schulabschluss*. Bei gestapelten Schichten entspricht die Gesamtzahl Schichten der Summe der Anzahl Kategorien aller Schichtvariablen (einschließlich der Kategorien für Gesamtsummen und Zwischenergebnisse, falls Sie diese für die Schichtvariablen angefordert haben).

Zwei verschachtelte kategoriale Schichtvariablen

Beim Verschachteln von kategorialen Schichtvariablen wird für jede Kombination von Kategorien der Schichtvariablen eine separate Schicht erstellt.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Wenn Sie dies nicht bereits getan haben, befolgen Sie die Anweisungen im vorherigen Beispiel zum Erstellen einer Tabelle mit gestapelten Schichten.
3. Wählen Sie im Gruppenfeld "Schichtausgabe" die Option **Jede Kombination von Kategorien als Schicht zeigen** aus. Dies entspricht der Verschachtelung.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.
5. Doppelklicken Sie im Fenster des Viewers auf die Tabelle, um diese zu aktivieren.
6. Klicken Sie auf den Abwärtspfeil der Dropdown-Liste, um die vollständige Liste der Schichten anzuzeigen.

Die Tabelle enthält 10 Schichten (Sie müssen in der Liste blättern, um alle Schichten anzuzeigen): eine Schicht für jede Kombination von *Geschlecht* und *Höchster Schulabschluss*. Bei verschachtelten Schichten entspricht die Gesamtzahl der Schichten dem *Produkt* aus den Anzahlen der Kategorien jeder Schichtvariablen (im vorliegenden Beispiel ist dies $5 \times 2 = 10$).

Drucken von Tabellen mit Schichten

In der Standardeinstellung wird nur die aktuell sichtbare Schicht gedruckt. So drucken Sie alle Schichten in einer Tabelle:

1. Doppelklicken Sie im Fenster des Viewers auf die Tabelle, um diese zu aktivieren.
2. Wählen Sie die folgenden Befehle aus den Menüs des Viewer-Fensters aus:
Format > Tabelleneigenschaften...
3. Klicken Sie auf die Registerkarte **Drucken**.
4. Wählen Sie **Alle Schichten drucken** aus.

Sie können diese Einstellung in einer Tabellenvorlage (auch der Standardtabellenvorlage) speichern.

Kapitel 4. Gesamtsummen und Zwischenergebnisse für kategoriale Variablen

Benutzerdefinierte Tabellen können sowohl Gesamtsummen als auch Zwischenergebnisse enthalten. Gesamtsummen und Zwischenergebnisse können kategorialen Variablen auf jeder Verschachtelungsebene und in jeder Dimension (Zeile, Spalte oder Schicht) zugewiesen werden.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Einfache Gesamtsumme für eine Variable

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus.
4. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
5. Löschen Sie in der Liste "Anzeigen" in der Zelle mit der Beschriftung die Standardbeschriftung und geben Sie Prozent ein.
6. Klicken Sie auf **Der Auswahl zuweisen**.
7. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
8. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken.
9. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Abbildung 20. Einfache Gesamtsumme für eine kategoriale Variable

Gesamtsumme entspricht der Summe aller angezeigten Werte

Die Gesamtsumme wird auf der Grundlage der in der Tabelle angezeigten Kategorien berechnet. Wenn Sie einige Kategorien aus der Tabelle ausschließen, werden die Fälle dieser Kategorien in der Berechnung der Gesamtsumme nicht berücksichtigt.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Klicken Sie in der Liste "Beschriftung" auf die Kategorie mit der Beschriftung *Unter 25*.
4. Klicken Sie auf die Pfeilschaltfläche links neben der Liste "Ausschließen".
5. Klicken Sie in der Liste "Beschriftung" auf die Kategorie mit der Beschriftung *65 oder älter*.
6. Klicken Sie erneut auf die Pfeilschaltfläche links neben der Liste "Ausschließen".
Daraufhin werden die beiden Kategorien aus der Liste "Anzeigen" in die Liste "Ausschließen" verschoben.
7. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count	Percent
Age category	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%
	Total	2107	100.0%

Abbildung 21. Gesamtsumme in einer Tabelle mit ausgeschlossenen Kategorien

Die Gesamtanzahl in dieser Tabelle beträgt nur 2.107, im Vergleich zu 2.828 bei Berücksichtigung aller Kategorien. Nur die in der Tabelle verwendeten Kategorien werden bei der Berechnung der Gesamtsumme berücksichtigt. (Die Gesamtsumme für die Prozentsätze beträgt weiterhin 100 %, da die Prozentsätze anhand der Gesamtanzahl der in der Tabelle verwendeten Fälle berechnet werden, nicht anhand der Gesamtanzahl der Fälle in der Datendatei.)

Anzeigeposition der Gesamtsummen

In der Standardeinstellung werden Gesamtsummen unter den Kategorien angezeigt, auf die sie sich beziehen. Die Gesamtsummen können jedoch auch über den jeweiligen Kategorien angezeigt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Wählen Sie im Gruppenfeld "Gesamtsummen und Zwischenergebnisse erscheinen" die Option **Oberhalb der Kategorien, für die sie gelten** aus.
4. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count	Percent
Age category	Total	2107	100.0%
	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%

Abbildung 22. Gesamtsumme wird über den entsprechenden Kategorien angezeigt

Gesamtsummen für verschachtelte Tabellen

Da Gesamtsummen kategorialen Variablen auf jeder Ebene der Verschachtelung zugewiesen werden können, können Sie Tabellen erstellen, die Gesamtsummen für Gruppen auf mehreren Verschachtelungsebenen enthalten.

Gesamtsummen für Gruppen

Gesamtsummen für kategoriale Variablen, die in anderen kategorialen Variablen verschachtelt sind, stellen Gesamtsummen für Gruppen dar.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Geschlecht* in den Erstellungsbereich und legen Sie sie links neben *Alterskategorie* ab.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
Bevor Sie die Tabelle erstellen, legen Sie erneut fest, dass die Gesamtsummen wieder unter den entsprechenden Kategorien angezeigt werden.
4. Wählen Sie im Gruppenfeld "Gesamtsummen und Zwischenergebnisse erscheinen" die Option **Unterhalb der Kategorien, für die sie gelten** aus.
5. Klicken Sie auf **Zuweisen**, um diese Einstellung zu speichern und zum Tabellenerstellungsprogramm zurückzukehren.
6. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

			Count	Percent	
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%

Abbildung 23. Gesamtsummen von Alterskategorien innerhalb der Geschlechtskategorien

In der Tabelle werden nun zwei Gesamtsummen für Gruppen angezeigt: eine für Männer und eine für Frauen.

Gesamtergebnisse

Bei Gesamtsummen von verschachtelten Variablen handelt es sich immer um Gesamtsummen für Gruppen, nicht um Gesamtergebnisse. Wenn Sie Gesamtsummen für die gesamte Tabelle erhalten möchten, können Sie festlegen, dass Gesamtsummen für die Variable auf der äußersten Verschachtelungsebene erstellt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Geschlecht* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken.
4. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

			Count	Percent	
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
		Total	942	100.0%	
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
		Total	1165	100.0%	
	Total	Age category	25 to 34	627	29.8%
			35 to 44	679	32.2%
			45 to 54	481	22.8%
			55 to 64	320	15.2%
Total		2107	100.0%		

Abbildung 24. Gesamtergebnisse für eine verschachtelte Tabelle

Beachten Sie, dass das Gesamtergebnis lediglich 2.107 beträgt, nicht 2.828. Zwei Alterskategorien sind noch immer aus der Tabelle ausgeschlossen und die Fälle in diesen Kategorien werden deshalb nicht bei den Gesamtsummen berücksichtigt.

Gesamtsummen für Schichtvariablen

Gesamtsummen für Schichtvariablen werden als separate Schichten in der Tabelle angezeigt.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie im Tabellenerstellungsprogramm auf **Schichten**, um die Liste "Schichten" anzuzeigen.
3. Ziehen Sie die Variable *Geschlecht* aus dem Bereich für die Zeilen im Erstellungsbereich in die Liste "Schichten".

Hinweis: Da Sie bereits festgelegt haben, dass für *Geschlecht* Gesamtsummen berechnet werden, müssen Sie dies jetzt nicht mehr tun. Das Verschieben der Variablen zwischen den Dimensionen hat keine Auswirkungen auf die Einstellungen für diese Variable.

4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.
5. Doppelklicken Sie im Viewer auf die Pivot-Tabelle, um diese zu aktivieren.
6. Klicken Sie in der Dropdown-Liste "Schicht" auf den Abwärtspfeil, um eine Liste aller Schichten in der Tabelle anzuzeigen.

Die Tabelle enthält drei Schichten: *Geschlecht männlich*, *Geschlecht weiblich* und *Geschlecht gesamt*.

Anzeigeoption der Schichtsummen

Bei Gesamtsummen für Schichtvariablen bestimmt die Anzeigeposition (über bzw. unter) der Gesamtsummen die Position der Schicht mit den Gesamtsummen. Wenn Sie beispielsweise für die Gesamtsumme einer Schichtvariablen die Option **Oberhalb der Kategorien, für die sie gelten** angeben, wird die Schicht mit den Gesamtsummen als erste Schicht angezeigt.

Zwischenergebnisse

Sie können auch Zwischenergebnisse für Subsets von Kategorien einer Variablen aufnehmen. Beispielsweise können Sie Zwischenergebnisse für Alterskategorien berechnen lassen, die alle Befragten der Beispielumfrage im Alter von unter und über 45 Jahren darstellt.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
4. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
5. Wählen Sie in der Liste "Wert(e)" den Eintrag **3,00** aus.
6. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
7. Geben Sie im Textfeld "Beschriftung" Zwischenergebnis < 45 ein.
8. Klicken Sie dann auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die ersten drei Alterskategorien ein.
9. Wählen Sie in der Liste "Wert(e)" den Eintrag **6,00** aus.
10. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
11. Geben Sie im Textfeld "Beschriftung" Zwischenergebnis 45+ ein.
12. Klicken Sie dann auf **Weiter**.
Wichtiger Hinweis: Bevor Sie Zwischenergebnisse definieren, sollten Sie die Anzeigeposition für die Gesamtsummen und Zwischenergebnisse (**Oberhalb der Kategorien, für die sie gelten** oder **Unterhalb der Kategorien, für die sie gelten**) festlegen. Wenn Sie die Anzeigeposition ändern, hat dies Auswirkungen auf alle Zwischenergebnisse (nicht nur auf das zurzeit ausgewählte). Außerdem werden *die in den Zwischenergebnissen enthaltenen Kategorien geändert*.
13. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Subtotal < 45	1548
	45 to 54	481
	55 to 64	320
	65 or older	479
	Subtotal 45+	1280

Abbildung 25. Zwischenergebnisse für die Alterskategorie

Zwischenergebnis entspricht der Summe der angezeigten Werte

Genau wie Gesamtsummen beruhen Zwischenergebnisse auf den in der Tabelle enthaltenen Kategorien.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
Hinweis: Der für das erste Zwischenergebnis angezeigte Wert (nicht die Beschriftung) beträgt **1,00 ... 3,00**. Dies bedeutet, dass das Zwischenergebnis alle Werte der Liste zwischen 1 und 3 einschließt.
3. Wählen Sie in der Liste "Wert(e)" den Eintrag **1,00** aus (oder klicken Sie auf die Beschriftung *Unter 25*).
4. Klicken Sie auf die Pfeilschaltfläche links neben der Liste "Ausschließen".

Die erste Alterskategorie ist nun ausgeschlossen und der für das erste Zwischenergebnis angezeigte Wert ändert sich in **2,00 ... 3,00**. Dies bedeutet, dass die ausgeschlossene Kategorie nicht im Zwischenergebnis enthalten sein wird, da Zwischenergebnisse auf den in der Tabelle enthaltenen Kategorien beruhen. Durch das Ausschließen einer Kategorie wird diese automatisch aus den Zwischenergebnissen ausgeschlossen, sodass Sie beispielsweise nicht die Zwischenergebnisse ohne die dazugehörigen Kategorien anzeigen lassen können.

Ausblenden der Kategorien für das Zwischenergebnis

Sie haben die Möglichkeit, die Anzeige der Kategorien, die zur Definition eines Zwischenergebnisses dienen, zu unterdrücken und nur das Zwischenergebnis selbst anzuzeigen, also die Anzeige von Kategorien ohne Auswirkungen auf die zugrunde liegenden Daten zu reduzieren.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
4. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Pop-up-Menü die Option **Kategorien und Gesamtsummen** aus.
5. Wählen Sie in der Liste "Wert(e)" den Eintrag **3,00** aus.
6. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
7. Geben Sie im Textfeld "Beschriftung" *Weniger als 45* ein.
8. Aktivieren Sie **Zwischenergebniskategorien aus Tabelle ausblenden**.
9. Klicken Sie dann auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die ersten drei Alterskategorien ein.
10. Wählen Sie in der Liste "Wert(e)" den Eintrag **6,00** aus.
11. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
12. Geben Sie im Textfeld "Beschriftung" *45 oder älter* ein.
13. Aktivieren Sie **Kategorien für Zwischenergebnis ausblenden**.
14. Klicken Sie dann auf **Weiter**.
15. Um eine Gesamtsumme mit den Zwischenergebnissen in die Tabelle aufzunehmen, aktivieren Sie **Gesamt** im Gruppenfeld "Anzeigen".
16. Klicken Sie auf **Zuweisen**.
Im Erstellungsbereich wird die Tatsache berücksichtigt, dass die Zwischenergebnisse angezeigt, die Kategorien, die zur Definition der Zwischenergebnisse dienen, jedoch ausgeschlossen werden.
17. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count
Age category	Less than 45	1548
	45 or older	1280
	Total	2828

Abbildung 26. Tabelle nur mit Zwischenergebnissen und Gesamtsummen

Zwischenergebnisse für Schichtvariablen

Genau wie Gesamtsummen werden Zwischenergebnisse für Schichtvariablen als separate Schichten in der Tabelle angezeigt. Im wesentlichen werden die Zwischenergebnisse wie Kategorien behandelt. Jede Kategorie stellt eine eigene Schicht in der Tabelle dar und die Anzeigereihenfolge der Kategorien in den Schichten wird durch die im Dialogfeld "Kategorien und Gesamtsummen" festgelegte Reihenfolge für die Kategorien bestimmt. Hierbei werden die Anzeigepositionen der Zwischenergebniskategorien berücksichtigt.

Kapitel 5. Berechnete Kategorien für kategoriale Variablen

Sie können berechnete Kategorien in benutzerdefinierte Tabellen aufnehmen. Hierbei handelt es sich um neue Kategorien, die aus Kategorien derselben Variable in einer beliebigen Verschachtelungsebene in einer beliebigen Dimension (Zeile, Spalte oder Schicht) berechnet werden. Sie könnten zum Beispiel eine berechnete Kategorie aufnehmen, die die Differenz zwischen zwei Kategorien anzeigt.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

Einfache berechnete Kategorie

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
4. Wählen Sie in der Liste "Wert(e)" den Eintrag **3,00** aus.
5. Klicken Sie auf **Kategorie hinzufügen**, um das Dialogfeld "Berechnete Kategorie definieren" anzuzeigen.
6. Geben Sie im Textfeld "Beschriftung für berechnete Kategorie" weniger als 45 ein.
7. Wählen Sie in der Liste "Kategorien" **Weniger als 25 (1,00)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren. [1] wird im Ausdruck angezeigt.
8. Klicken Sie im Dialogfeld auf die Schaltfläche für den Plusoperator (+) oder drücken Sie die Plusstaste auf der Tastatur).
9. Wählen Sie in der Liste "Kategorien" **25 bis 34 (2,00)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
10. Klicken Sie im Dialogfeld auf die Schaltfläche für den Plusoperator (+) oder drücken Sie die Plusstaste auf der Tastatur).
11. Wählen Sie in der Liste "Kategorien" **35 bis 44 (3,00)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
12. Klicken Sie dann auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die ersten drei Alterskategorien ein.
13. Wählen Sie in der Liste "Wert(e)" den Eintrag **5,00** aus.
14. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
15. Geben Sie im Textfeld "Beschriftung" weniger als 65 ein.
16. Klicken Sie dann auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die ersten fünf Kategorien ein.
17. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Less than 45	1548
	45 to 54	481
	55 to 64	320
	Less than 65	2349
	65 or older	479

Abbildung 27. Berechnete Kategorie mit Zwischenergebnis

Die Tabelle enthält eine berechnete Kategorie (*Weniger als 45*) und ein Zwischenergebnis (*Weniger als 65*). Das Zwischenergebnis enthält Kategorien, die auch in der berechneten Kategorie enthalten sind. Sie könnten dieselbe Tabelle nur mit den Zwischenergebnissen nicht erstellen, da Zwischenergebnisse nicht über dieselben Kategorien verfügen.

Ausblenden von Kategorien in einer berechneten Kategorie

Wie bei den Zwischenergebnissen können Sie auch die Anzeige der Kategorien ausblenden, die im Ausdruck einer berechneten Kategorie verwendet werden, und nur die berechnete Kategorie selbst anzeigen. Das folgende Beispiel baut auf dem vorherigen auf.

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Wählen Sie die berechnete Kategorie *Weniger als 45* in der Liste "Wert(e)" aus.
4. Klicken Sie auf **Bearbeiten**, um das Dialogfeld "Berechnete Kategorie definieren" anzuzeigen.
5. Wählen Sie **Im Ausdruck verwendete Kategorien aus Tabelle ausblenden** aus.
6. Klicken Sie dann auf **Weiter**.
7. Wählen Sie das Zwischenergebnis *Weniger als 65* in der Liste "Wert(e)" aus.
8. Klicken Sie auf **Bearbeiten**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
9. Wählen Sie **Zwischenergebniskategorien aus Tabelle ausblenden** aus.
10. Klicken Sie dann auf **Weiter**.
11. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count
Age category	Less than 45	1548
	Less than 65	2349
	65 or older	479

Abbildung 28. Berechnete Kategorie mit Zwischenergebnis und ausgeblendeten Kategorien

Wie im vorherigen Beispiel enthält die Tabelle eine berechnete Kategorie und ein Zwischenergebnis. Doch in diesem Fall sind die Kategorien ausgeblendet und es werden nur die Zwischenergebnisse angezeigt.

Verweis auf Zwischenergebnisse in einer berechneten Kategorie

Sie können Zwischenergebnisse in den Ausdruck einer berechneten Kategorie aufnehmen.

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen

2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Beschäftigungsstatus* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich für die Zeilen im Erstellungsbereich.
4. Ziehen Sie die Variable *Familienstand* aus der Variablenliste und legen Sie sie im Spaltenbereich ab.
5. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Beschäftigungsstatus* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
6. Wählen Sie in der Liste "Wert(e)" den Eintrag **2** aus.
7. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
8. Geben Sie im Textfeld "Beschriftung" *Beschäftigt* ein.
9. Wählen Sie **Zwischenergebniskategorien aus Tabelle ausblenden** aus.
10. Klicken Sie dann auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die ersten zwei Kategorien des Beschäftigungsstatus ein.
11. Wählen Sie in der Liste "Wert(e)" den Eintrag **8** aus.
12. Klicken Sie auf **Zwischenergebnis hinzufügen**, um das Dialogfeld "Zwischenergebnis definieren" anzuzeigen.
13. Geben Sie im Textfeld "Beschriftung" *Nicht beschäftigt* ein.
14. Wählen Sie **Kategorien für Zwischenergebnis ausblenden** aus.
15. Klicken Sie dann auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die restlichen Kategorien des Beschäftigungsstatus ein.
16. Wählen Sie das Zwischenergebnis *Nicht beschäftigt* in der Liste "Wert(e)" aus.
17. Klicken Sie auf **Kategorie hinzufügen**, um das Dialogfeld "Berechnete Kategorie definieren" anzuzeigen.
18. Geben Sie im Textfeld "Beschriftung für berechnete Kategorie" *Beschäftigt / Nicht beschäftigt* ein.
19. Wählen Sie in der Liste "Gesamtsummen und Zwischenergebnisse" **Beschäftigt (Beschäftigt #1)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
20. Klicken Sie im Dialogfeld auf die Schaltfläche für den Divisionsoperator (/) oder drücken Sie die Taste für diesen Operator auf der Tastatur.
21. Wählen Sie in der Liste "Gesamtsummen und Zwischenergebnisse" **Nicht beschäftigt (Nicht berechnet #2)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
Die berechnete Kategorie verwendet standardmäßig dasselbe Format wie die Variablenstatistik, in diesem Fall "Anzahl". Da nach der Division im Ausdruck der berechneten Kategorie Dezimalstellen angezeigt werden sollen und das Standardformat für "Anzahl" keine Dezimalstellen enthält, muss das Format geändert werden.
22. Klicken Sie auf die Registerkarte "Anzeigeformate".
23. Ändern Sie die Einstellung der Dezimalstellen für "Anzahl" auf **2**.
24. Klicken Sie dann auf **Weiter**.
25. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Count	Count	Count	Count	Count
Labor force status	Working	916	64	330	67	494
	Not Working	429	219	116	26	169
	Working / Not Working	2.14	.29	2.84	2.58	2.92

Abbildung 29. Anzeige des Anteils von Zwischenergebnissen in berechneten Kategorien

Die Tabelle enthält zwei Zwischenergebnisse und eine berechnete Kategorie. Die berechnete Kategorie zeigt den Anteil der Zwischenergebnisse an, damit Sie die durch jedes Zwischenergebnis dargestellten Gruppen leicht vergleichen können. Der Anteil von beschäftigten und nicht beschäftigten verwitweten Teilnehmern ist deutlich niedriger als bei den anderen Gruppen. Außerdem herrscht ein etwas niedrigerer Anteil an verheirateten Teilnehmern, möglicherweise aufgrund von Lebensgefährten, die das Beschäftigungsverhältnis kündigen, um sich zu Hause um das Kind zu kümmern.

Einsatz von berechneten Kategorien zur Anzeige nicht abschließender Zwischenergebnisse

Zwischenergebnisse sind abschließend. Das heißt, alle Zwischenergebnisse in einer Tabelle enthalten alle Werte über und unter ihren Positionen in der Tabelle. Berechnete Kategorien sind dagegen nicht abschließend und ermöglichen Ihnen, eine Summe aus einer Mischung von Kategorien in der Tabelle zu erstellen.

- Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
- Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
- Ziehen Sie die Variable *Betrachtet sich selbst als liberal oder konservativ* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich für die Zeilen im Erstellungsbereich.
- Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Betrachtet sich selbst als liberal oder konservativ* und wählen Sie im Kontextmenü die Option **Kategorien und Gesamtsummen** aus.
- Wählen Sie in der Liste "Wert(e)" den Eintrag **3** aus.
- Klicken Sie auf **Kategorie hinzufügen**, um das Dialogfeld "Berechnete Kategorie definieren" anzuzeigen.
- Geben Sie im Textfeld "Beschriftung für berechnete Kategorie" **Liberales Zwischenergebnis** ein. Beachten Sie, dass vor dem Text vier Leerzeichen stehen. Diese vier Leerzeichen werden als Einrückung in der Endtabelle verwendet.
- Wählen Sie in der Liste "Kategorien" **Extrem liberal (1)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
- Klicken Sie im Dialogfeld auf die Schaltfläche für den Plusoperator (+) oder drücken Sie die Plustaste auf der Tastatur).
- Wählen Sie in der Liste "Kategorien" **Liberal (2)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
- Klicken Sie im Dialogfeld auf die Schaltfläche für den Plusoperator (+) oder drücken Sie die Plustaste auf der Tastatur).
- Wählen Sie in der Liste "Kategorien" **Leicht liberal (3)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
- Klicken Sie auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die liberalen Kategorien ein.
- Wählen Sie in der Liste "Wert(e)" den Eintrag **7** aus.
- Klicken Sie auf **Kategorie hinzufügen**, um das Dialogfeld "Berechnete Kategorie definieren" anzuzeigen.

16. Geben Sie im Textfeld "Beschriftung für berechnete Kategorie" Konservatives Zwischenergebnis ein. Beachten Sie, dass vor dem Text vier Leerzeichen stehen. Diese vier Leerzeichen werden als Einrückung in der Endtabelle verwendet.
17. Wählen Sie in der Liste "Kategorien" **Leicht konservativ (5)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
18. Klicken Sie im Dialogfeld auf die Schaltfläche für den Plusoperator (+) oder drücken Sie die Plusstaste auf der Tastatur).
19. Wählen Sie in der Liste "Kategorien" **Konservativ (6)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
20. Klicken Sie im Dialogfeld auf die Schaltfläche für den Plusoperator (+) oder drücken Sie die Plusstaste auf der Tastatur).
21. Wählen Sie in der Liste "Kategorien" **Extrem konservativ (7)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in das Textfeld "Ausdruck für berechnete Kategorie" zu kopieren.
22. Klicken Sie auf **Weiter**.
Hiermit fügen Sie eine Zeile mit dem Zwischenergebnis für die konservativen Kategorien ein.
23. Klicken Sie auf **Zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count
Think of self as liberal or conservative	Extremely liberal	64
	Liberal	357
	Slightly liberal	351
	Liberal Subtotal	772
	Moderate	986
	Slightly conservative	432
	Conservative	415
	Extremely conservative	86
	Conservative Subtotal	933

Abbildung 30. Anzeige von nicht abschließenden Zwischenergebnissen für berechnete Kategorien

Die Tabelle enthält zwei berechnete Kategorien, die nicht alle in der Tabelle angezeigten Kategorien enthalten. Die Kategorie *Mäßig* ist in keiner der berechneten Kategorien enthalten. Sie können dieselbe Tabelle mit Zwischenergebnissen nicht erstellen, da Zwischenergebnisse abschließend sind.

Kapitel 6. Tabellen für Variablen mit gemeinsamen Kategorien

Oft enthalten Umfragen viele Fragen mit übereinstimmenden möglichen Antworten. Die Beispielumfrage enthält beispielsweise eine Reihe von Variablen, die sich auf das Vertrauen in verschiedene öffentliche und private Institutionen und Dienstleistungen beziehen. Für diese Variablen ist jeweils dieselbe Gruppe von Antwortkategorien verfügbar: 1 = *Sehr viel*, 2 = *Mäßig* und 3 = *Fast kein*. Sie können diese zusammenhängenden Variablen stapeln, damit sie in einer Tabelle dargestellt werden, und die gemeinsamen Antwortkategorien können als Spalten der Tabelle erscheinen. Diese Funktionen sind auch verfügbar, wenn Sie berechnete Kategorien verwenden, wobei Voraussetzung ist, dass alle Beschriftungen und Ausdrücke berechneter Kategorien in sämtlichen Variablen übereinstimmen.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Abbildung 31. Tabelle für Variablen mit gemeinsamen Kategorien

Hinweis: In der Vorversion der Funktion für benutzerdefinierte Tabellen wurden diese als "Häufigkeitstabellen" bezeichnet.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Häufigkeitstabelle

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

2. Klicken Sie in der Variablenliste des Tabellenerstellungsprogramms auf *Vertrauen in Banken...* und klicken Sie anschließend mit gedrückter Umschalttaste auf *Vertrauen in das Fernsehen*, um alle Vertrauensvariablen auszuwählen. (*Hinweis:* Hierbei wird vorausgesetzt, dass die Variablenbeschriftungen in der Variablenliste in alphabetischer Reihenfolge und nicht in der Reihenfolge angezeigt werden, in der sich in der Datei befinden.)

3. Ziehen Sie die sechs Vertrauensvariablen in den Erstellungsbereich und legen Sie sie im Bereich für die Zeilen ab.

Hiermit werden die Variablen in der Zeilendimension gestapelt. In der Standardeinstellung werden die Kategoriebeschriftungen für jede Variable auch in den Zeilen angezeigt, wodurch eine sehr lange, schmale Tabelle entsteht (6 Variablen x 3 Kategorien = 18 Zeilen). Da jedoch für alle sechs Variablen dieselben Kategoriebeschriftungen (Wertbeschriftungen) definiert sind, können Sie die Kategoriebeschriftungen in die Spaltendimension aufnehmen.

- Wählen Sie in der Dropdown-Liste "Kategorieposition" die Option **Zeilenbeschriftungen in Spalten** aus.
Nun besitzt die Tabelle nur sechs Zeilen für jeweils eine der gestapelten Variablen und die definierten Kategorien erscheinen als Spalten der Tabelle.
- Bevor Sie die Tabelle erstellen, aktivieren Sie im Gruppenfeld "Auswertungsstatistik" neben der Dropdown-Liste "Position" die Option **Verbergen**, indem Sie darauf klicken, da die Beschriftung für die Auswertungsstatistik *Anzahl* nicht unbedingt erforderlich ist.
- Klicken Sie auf **OK**, um die Tabelle zu erstellen.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Abbildung 32. Tabelle für gestapelte Zeilenvariablen mit gemeinsamen Kategoriebeschriftungen in Spalten

Die Variablen müssen nicht notwendigerweise in Zeilen und die Kategorien in Spalten angezeigt werden. Sie können auch eine Tabelle erstellen, in der die Variablen in den Spalten gestapelt und die Kategorien in den Zeilen angezeigt werden. Dies ist vorteilhaft, wenn mehr Kategorien als Variablen vorhanden sind. Im vorliegenden Beispiel gibt es jedoch mehr Variablen als Kategorien.

Tabelle mit Prozentsätzen

Bei Tabellen, in denen die Variablen in Zeilen gestapelt und die Kategorien in Spalten angezeigt werden, stellen Zeilenprozent die aussagekräftigsten (oder zumindest die am leichtesten verständlichen) Prozentsätze dar. (Bei Tabellen, in denen die Variablen in den Spalten gestapelt und die Kategorien in den Zeilen angezeigt werden, würden Sie eher Spaltenprozent wählen.)

- Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
- Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf eine der Vertrauensvariablen und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
- Wählen Sie in der Liste "Statistik" die Option **Anzahl als Zeilen%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
- Klicken Sie auf eine beliebige Zelle in der Zeile *Anzahl* und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Statistik" zu übertragen und aus der Liste "Anzeigen" zu entfernen.
- Klicken Sie auf **Allen zuweisen**, um die Änderung der Auswertungsstatistik allen gestapelten Variablen in der Tabelle zuzuweisen.

Hinweis: Wenn die Tabellenvorschau nicht dieser Abbildung entspricht, haben Sie vermutlich nicht auf **Allen zuweisen**, sondern auf **Der Auswahl zuweisen** geklickt. Hierbei wird die neue Auswertungsstatistik nur der ausgewählten Variablen zugewiesen. In diesem Beispiel würden dann zwei Spalten für jede Kategorie entstehen: eine Spalte mit Platzhaltern für die Anzahl, die für alle anderen Variablen gilt, und eine Spalte mit einem Platzhalter für Zeilenprozent, die für die ausgewählte Variable gilt. Diese Tabelle würde dann erzeugt werden, jedoch ist dies im vorliegenden Beispiel *nicht* erwünscht.

- Klicken Sie auf **OK**, um die Tabelle zu erstellen.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	26.3%	57.3%	16.4%
Confidence in education	27.2%	56.1%	16.7%
Confidence in major companies	27.5%	59.2%	13.3%
Confidence in medicine	45.0%	46.1%	8.9%
Confidence in press	9.5%	47.2%	43.4%
Confidence in television	10.4%	49.9%	39.7%

Abbildung 33. Tabelle mit Zeilenprozenten für die in Zeilen gestapelten Variablen; die Kategorien werden in Spalten angezeigt

Hinweis: Eine Tabelle für Variablen mit gemeinsamen Kategorien kann eine beliebige Anzahl von Auswertungsstatistiken enthalten. In den vorliegenden Beispielen wird der Einfachheit halber jeweils nur eine dargestellt.

Gesamtsummen und Kategorien

Tabellen, in denen sich die Kategorien in der entgegengesetzten Dimension der Variablen befinden, können nur dann erstellt werden, wenn alle Variablen in der Tabelle dieselben Kategorien aufweisen und diese in derselben Reihenfolge angezeigt werden. Dies betrifft Gesamtsummen, Zwischenergebnisse und alle weiteren Anpassungen an den Kategorien. Dies bedeutet, dass Sie sämtliche Änderungen im Dialogfeld "Kategorien und Gesamtsummen" für alle Variablen der Tabelle vornehmen müssen, die dieselben Kategorien aufweisen.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die erste Vertrauensvariable und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Aktivieren Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt**, indem Sie darauf klicken, und klicken Sie anschließend auf **Zuweisen**.

Als erstes wird Ihnen wahrscheinlich auffallen, dass die Kategoriebeschriftungen aus den Spalten wieder in die Zeilen verschoben wurden. Sie werden vielleicht auch bemerken, dass die Dropdown-Liste "Kategorieposition" nun inaktiviert ist. Die Ursache hierfür ist, dass die Variablen nicht mehr genau dieselbe Gruppe von "Kategorien" aufweisen. Eine der Variablen besitzt nun eine Kategorie für Gesamtsummen.

4. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf eine beliebige Vertrauensvariable und wählen Sie im Popup-Menü die Option **Alle Zeilenvariablen auswählen**. Sie können auch bei gedrückter Steuertaste im Erstellungsbereich auf jede gestapelte Variable klicken, bis alle Variablen ausgewählt sind (möglicherweise müssen Sie nach unten blättern oder das Fenster des Tabellenerstellungsprogramms vergrößern).
5. Klicken Sie im Gruppenfeld "Definieren" auf **Kategorien und Gesamtsummen**.
6. Wenn im Dialogfeld "Kategorien und Gesamtwerte" die Option **Gesamt** nicht bereits ausgewählt (aktiviert) ist, wählen Sie diese Option jetzt aus und klicken Sie dann auf **Zuweisen**.
7. Die Dropdown-Liste "Kategorieposition" ist nun wieder aktiviert, da jetzt alle Variablen über die zusätzliche Kategorie für Gesamtsummen verfügen. Wählen Sie daher **Zeilenbeschriftungen in Spalten** aus.
8. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

	A great deal	Only some	Hardly any	Total
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%
Confidence in television	10.4%	49.9%	39.7%	100.0%

Abbildung 34. Tabelle mit Zeilenprozenten für die in Zeilen gestapelten Variablen; die Kategorien und Gesamtsummen werden in Spalten angezeigt

Verschachteln in Tabellen mit gemeinsamen Kategorien

In verschachtelten Tabellen müssen sich die gestapelten Variablen mit den gemeinsamen Kategorien auf der innersten Verschachtelungsebene der betreffenden Dimension befinden, wenn die Kategoriebeschriftungen in der entgegengesetzten Dimension angezeigt werden sollen.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste und legen Sie sie an der linken Seite des Zeilenbereichs ab.

In der Tabellenvorschau sind die gestapelten Variablen mit gemeinsamen Kategorien jetzt innerhalb der Geschlechtskategorien verschachtelt.

3. Ziehen Sie nun die Variable *Geschlecht* rechts neben eine der gestapelten Vertrauensvariablen in der Tabellenvorschau und legen Sie sie dort ab.

Die Kategoriebeschriftungen wurden wieder in die Zeilendimension verschoben und die Dropdown-Liste "Kategorieposition" ist inaktiviert. Es ist nun eine gestapelte Variable vorhanden, in der auch die Kategorie *Geschlecht* verschachtelt ist. Die anderen gestapelten Variablen enthalten hingegen keine verschachtelten Variablen. Sie könnten zwar jeder der gestapelten Variablen die Variable *Geschlecht* als verschachtelte Variable hinzufügen. Beim Verschieben der Zeilenbeschriftungen in die Spalten würden dann jedoch nicht die Kategoriebeschriftungen für die gestapelten Variablen mit den gemeinsamen Kategorien, sondern die Kategoriebeschriftungen für *Geschlecht* in den Spalten angezeigt werden. Der Grund hierfür besteht darin, dass *Geschlecht* nun die innerste verschachtelte Variable bilden würde und sich Änderungen an der Kategorieposition immer auf die innerste verschachtelte Variable auswirken.

Kapitel 7. Auswertungsstatistik

Zu den **Auswertungsstatistiken** gehören verschiedene Statistiken, von einfachen Häufigkeiten kategorialer Variablen bis hin zu Streuungsmaßen, z. B. der Standardfehler des Mittelwerts für metrische Variablen. *Nicht* hierzu zählen jedoch Signifikanztests, die im Dialogfeld "Benutzerdefinierte Tabellen" auf der Registerkarte "Teststatistiken" verfügbar sind. Weitere Informationen finden Sie in „Teststatistiken“ auf Seite 73.

Als Auswertungsstatistiken für kategoriale Variablen und Mehrfachantwortsets sind Häufigkeiten und eine breite Auswahl von Prozentrechnungen verfügbar. Dazu gehören:

- Zeilenprozente
- Spaltenprozente
- Untertabellenprozente
- Tabellenprozente
- Gültige N als Prozente

Neben den Auswertungsstatistiken für kategoriale Variablen sind die folgenden Auswertungsstatistiken für metrische Variablen und benutzerdefinierten Gesamtstatistiken für kategoriale Variablen verfügbar:

- Mittelwert
- Median
- Perzentile
- Summe
- Standardabweichung
- Spannweite
- Minimal- und Maximalwerte

Für Mehrfachantwortsets sind weitere Auswertungsstatistiken verfügbar. Außerdem ist eine vollständige Liste von Auswertungsstatistiken verfügbar. Weitere Informationen finden Sie in „Auswertungsstatistik“ auf Seite 6.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Quellenvariable für Auswertungsstatistik

Es hängt vom Messniveau der jeweiligen Quellenvariablen ab, welche Auswertungsstatistiken verfügbar sind. Die Quelle der Auswertungsstatistiken (die Variable, auf der die Auswertungsstatistiken beruhen) wird durch die folgenden Faktoren bestimmt:

- **Messniveau.** Wenn eine Tabelle (oder ein Tabellenabschnitt in einer gestapelten Tabelle) eine metrische Variable enthält, beruhen Auswertungsstatistiken auf der metrischen Variablen.
- **Reihenfolge der Variablenauswahl.** Die Standardquellendimension der Statistik (Zeile oder Spalte) für kategoriale Variablen beruht auf der Reihenfolge, in der Sie die Variablen in den Erstellungsbereich zie-

hen. Wenn Sie z. B. eine Variable zuerst in den Bereich für die Zeilen ziehen, wird die Zeilendimension als Standardquellendimension für die Statistik verwendet.

- **Verschachtelung.** Bei kategorialen Variablen beruhen Auswertungsstatistiken auf der innersten Variablen in der Quellendimension der Statistik.

Eine gestapelte Tabelle kann mehrere Quellenvariablen für Auswertungsstatistiken aufweisen (sowohl metrisch als auch kategorial); jeder Tabellenabschnitt besitzt jedoch nur eine Quelle für die Auswertungsstatistik.

Quelle von Auswertungsstatistiken für kategoriale Variablen

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich für die Zeilen im Erstellungsbereich.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus. (Da es sich um die einzige Variable in der Tabelle handelt, wird sie als deren Quellenvariable verwendet.)
4. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option *Anzahl als Spalten%* aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
5. Klicken Sie auf **Der Auswahl zuweisen**.
6. Ziehen Sie im Tabellenerstellungsprogramm die Variable *Nachrichten aus dem Internet* in den Erstellungsbereich und legen Sie sie rechts neben *Alterskategorie* ab.
7. Klicken Sie mit der rechten Maustaste erneut im Erstellungsbereich auf *Alterskategorie*. Die Option **Auswertungsstatistik** im Popup-Menü ist nun inaktiviert, da *Alterskategorie* nicht die innerste Variable in der Quellendimension der Statistik ist.
8. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Nachrichten aus dem Internet*. Die Option **Auswertungsstatistik** ist aktiviert, weil diese Variable nun die innerste verschachtelte Variable der Quellendimension der Statistik ist und somit die Quellenvariable der Auswertungsstatistik darstellt. (Da die Tabelle nur eine Dimension (Zeilen) besitzt, ist diese die Quellendimension der Statistik.)
9. Ziehen Sie die Variable *Nachrichten aus dem Internet* im Erstellungsbereich aus dem Bereich für die Zeilen und legen Sie sie im Bereich für die Spalten ab.
10. Klicken Sie mit der rechten Maustaste im Erstellungsbereich erneut auf *Nachrichten aus dem Internet*. Die Option **Auswertungsstatistik** im Popup-Menü ist nun inaktiviert, da sich die Variable nicht mehr in der Quellendimension der Statistik befindet.

Jetzt ist *Alterskategorie* wieder die Quellenvariable der Statistik, da die Standardquellendimension der Statistik für kategoriale Variablen die erste Dimension ist, in der Sie beim Erstellen der Tabelle Variablen ablegen. In diesem Beispiel wurden zuerst Variablen in der Zeilendimension abgelegt. Die Zeilendimension ist daher die Standardquellendimension der Statistik. Da *Alterskategorie* nun die einzige Variable in dieser Dimension ist, wird sie als Quellenvariable der Statistik verwendet.

Quelle von Auswertungsstatistiken für metrische Variablen

1. Ziehen Sie die metrische Variable *Täglicher Fernsehkonsum in Stunden* aus dem Zeilenbereich des Erstellungsbereichs und legen Sie sie links neben der Variablen *Alterskategorie* ab.
Zunächst ist erkennbar, dass die Auswertungen *Anzahl* und *Anzahl als Spalten%* durch *Mittelwert* ersetzt wurden. Wenn Sie mit der rechten Maustaste im Erstellungsbereich auf *Täglicher Fernsehkonsum in Stunden* klicken, wird sichtbar, dass diese Variable nun als Quellenvariable der Auswertungsstatistik verwendet wird. In Tabellen mit einer metrischen Variablen wird diese stets als Quellenvariable der Statistik verwendet. Dies gilt unabhängig von der Verschachtelungsebene oder Dimension und als Standardauswertungsstatistik für metrische Variablen dient der Mittelwert.

- Ziehen Sie die Variable *Täglicher Fernsehkonsum in Stunden* aus dem Zeilenbereich und legen Sie sie im Spaltenbereich über der Variablen *Nachrichten aus dem Internet* ab.
- Klicken Sie mit der rechten Maustaste auf *Täglicher Fernsehkonsum in Stunden* und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus. (Diese bleibt als Quellenvariable der Statistik erhalten, auch wenn Sie sie in eine andere Dimension verschieben.)
- Klicken Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Anzeigen" auf die Zelle **Format** für den Mittelwert und wählen Sie aus der Dropdown-Liste "Format" den Eintrag **nnnn** aus. (Möglicherweise müssen Sie in der Liste blättern, um diese Option zu finden.)
- Geben Sie in der Zelle "Dezimalstellen" den Wert 2 ein.
- Klicken Sie auf **Der Auswahl zuweisen**.
Die Tabellenvorschau im Erstellungsbereich verdeutlicht nun, dass die Mittelwerte mit zwei Dezimalstellen angezeigt werden.
- Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Hours per day watching TV	
		Get news from internet	
		No	Yes
		Mean	Mean
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

Abbildung 35. Auswertung einer metrischen Variablen innerhalb von kategorialen Variablen in einer Kreuztabelle

Gestapelte Variablen

Da eine gestapelte Tabelle mehrere Quellenvariablen für Statistiken enthalten kann und Sie für diese Quellenvariablen unterschiedliche Auswertungsstatistiken festlegen können, müssen beim Festlegen von Auswertungsstatistiken in gestapelten Tabellen einige besondere Überlegungen getroffen werden.

- Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
- Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
- Klicken Sie in der Variablenliste auf *Nachrichten aus dem Internet* und klicken Sie bei gedrückter Umschalttaste in der Variablenliste auf *Nachrichten aus dem Fernsehen*, um alle Nachrichtenvariablen auszuwählen. (*Hinweis*: Hierbei wird vorausgesetzt, dass die Variablenbeschriftungen in der Variablenliste in alphabetischer Reihenfolge und nicht in der Reihenfolge angezeigt werden, in der sich in der Datei befinden.)
- Ziehen Sie die fünf Nachrichtenvariablen in den Bereich für die Zeilen im Erstellungsbereich und legen Sie sie dort ab.
Die fünf Nachrichtenvariablen werden in der Zeilendimension gestapelt.
- Klicken Sie im Erstellungsbereich auf *Nachrichten aus dem Internet*, sodass nur diese Variable ausgewählt ist.
- Klicken Sie jetzt mit der rechten Maustaste auf *Nachrichten aus dem Internet* und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus.
- Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option *Anzahl als Spalten%* aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen. (Mit der Pfeilschaltfläche können Sie ausgewählte Statistiken aus der Liste "Statistik" in die Liste "Anzeigen" übertragen. Sie können die Statistiken auch aus der Liste "Statistik" in die Liste "Anzeigen" ziehen und dort ablegen.)

8. Klicken Sie dann auf **Der Auswahl zuweisen**.
Eine Spalte für Spaltenprozentage wird hinzugefügt. In der Tabellenvorschau im Erstellungsbereich wird jedoch angezeigt, dass Spaltenprozentage nur für eine Variable angezeigt werden. Der Grund dafür besteht darin, dass in einer gestapelten Tabelle mehrere Quellenvariablen für Statistiken vorhanden sind und für jede dieser Variablen eine andere Auswertungsstatistik vorhanden sein kann. In diesem Beispiel soll jedoch für alle Variablen dieselbe Auswertungsstatistik angezeigt werden.
9. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf *Nachrichten aus Zeitungen* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
10. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option *Anzahl als Spalten%* aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
11. Klicken Sie dann auf **Allen zuweisen**.

In der Tabellenvorschau wird nun dargestellt, dass Spaltenprozentage für alle gestapelten Variablen angezeigt werden.

Angepasste Auswertungsstatistik für Gesamtsummen bei kategorialen Variablen

Bei kategorialen Quellenvariablen für Statistiken können Sie angepasste Auswertungsstatistiken für Gesamtsummen einbeziehen, die sich von den Statistiken unterscheiden, die für die Kategorien der Variablen gezeigt werden. Bei einer ordinalen Variablen können Sie beispielsweise Prozente für jede Kategorie sowie den Mittelwert oder Median als angepasste Auswertungsstatistik für Gesamtsummen anzeigen.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Klicken Sie in der Variablenliste auf *Vertrauen in die Presse* und klicken Sie dann bei gedrückter Steuertaste auf *Vertrauen in das Fernsehen*, um beide Variablen auszuwählen.
4. Ziehen Sie die beiden Variablen in den Bereich für die Zeilen im Erstellungsbereich und legen Sie sie dort ab. Hiermit werden die beiden Variablen in der Zeilendimension gestapelt.
5. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf eine der Variablen und wählen Sie im Popup-Menü die Option **Alle Zeilenvariablen auswählen** aus. (Dies ist nur als Absicherung gedacht; möglicherweise sind die Variablen bereits ausgewählt.)
6. Klicken Sie mit der rechten Maustaste erneut auf die Variable und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
7. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus (aktivieren Sie diese) und klicken Sie anschließend auf **Zuweisen**.
In der Tabellenvorschau im Erstellungsbereich wird nun eine Zeile für die Gesamtsumme für beide Variablen angezeigt. Um eine angepasste Auswertungsstatistik für die Gesamtsumme anzuzeigen, müssen für die Tabelle Gesamtsummen bzw. Zwischenergebnisse festgelegt werden.
8. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf eine der Variablen und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
9. Klicken Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Anzeigen" auf die Option *Anzahl* und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Statistik" zu übertragen und aus der Liste "Anzeigen" zu entfernen.
10. Klicken Sie in der Liste "Statistik" auf *Anzahl als Spalten%* und klicken Sie auf die Pfeilschaltfläche, um die Variable in die Liste "Anzeigen" zu verschieben.
11. Klicken Sie auf die Option **Angepasste Auswertungsstatistik für Gesamtsummen und Zwischenergebnisse** (aktivieren Sie diese).

12. Klicken Sie in der Liste "Anzeigen" der benutzerdefinierten Auswertung auf die Option *Anzahl* und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Statistik" der benutzerdefinierten Auswertung zu übertragen und aus der Liste "Anzeigen" zu entfernen.
13. Klicken Sie in der Liste "Statistik" der benutzerdefinierten Auswertung auf die Option *Mittelwert* und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" der benutzerdefinierten Auswertung zu übertragen.
14. Klicken Sie in der Liste "Anzeigen" auf die Zelle **Format** für den Mittelwert und wählen Sie aus der Dropdown-Liste der Formate den Eintrag **nnnn** aus. (Möglicherweise müssen Sie in der Liste blättern, um diese Option zu finden.)
15. Geben Sie in der Zelle "Dezimalstellen" den Wert 2 ein.
16. Klicken Sie auf **Allen zuweisen**, um diese Einstellungen beiden Variablen in der Tabelle zuzuweisen. Für die angepasste Auswertungsstatistik für Gesamtsummen wurde eine neue Spalte hinzugefügt. Dies ist jedoch möglicherweise unerwünscht, da in der Vorschau im Erstellungsbereich deutlich gezeigt wird, dass die Tabelle hierbei eine große Anzahl leerer Zellen enthält.
17. Wählen Sie im Tabellenerstellungsprogramm im Gruppenfeld "Auswertungsstatistik" den Eintrag **Zeilen** aus der Dropdown-Liste "Position " aus. Hiermit verschieben Sie alle Auswertungsstatistiken in die Zeilendimension, wodurch diese in einer einzigen Spalte der Tabelle angezeigt werden.
18. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

Confidence in press	A great deal	Column N %	9.5%
	Only some	Column N %	47.2%
	Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	A great deal	Column N %	10.4%
	Only some	Column N %	49.9%
	Hardly any	Column N %	39.7%
	Total	Mean	2.29

Abbildung 36. Kategoriale Variablen mit angepasster Auswertungsstatistik für Gesamtsummen

Anzeigen von Kategoriewerten

Im Zusammenhang mit der oben beschriebenen Tabelle besteht nur ein kleines Problem: Es kann schwierig sein, den Mittelwert zu interpretieren, ohne die zugrunde liegenden Kategoriewerte zu kennen. Befindet sich ein Mittelwert von 2,34 zwischen *Sehr viel* und *Mäßig* oder zwischen *Mäßig* und *Fast kein*?

Dieses Problem kann zwar nicht direkt für die benutzerdefinierte Tabelle, aber in allgemeinerer Form gelöst werden.

1. Wählen Sie in den Menüs Folgendes aus:
Bearbeiten > Optionen...
2. Klicken Sie im Dialogfeld "Optionen" auf die Registerkarte **Beschriftung der Ausgabe**.
3. Wählen Sie im Gruppenfeld "Beschriftung für Pivot-Tabellen" aus der Dropdown-Liste **Variablenwerte in Beschriftungen anzeigen als** die Option **Werte und Beschriftungen** aus.
4. Klicken Sie auf **OK**, um diese Einstellung zu speichern.
5. Öffnen Sie erneut das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen") und klicken Sie auf **OK**, um die Tabelle erneut zu erstellen.

Confidence in press	1 A great deal	Column N %	9.5%
	2 Only some	Column N %	47.2%
	3 Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	1 A great deal	Column N %	10.4%
	2 Only some	Column N %	49.9%
	3 Hardly any	Column N %	39.7%
	Total	Mean	2.29

Abbildung 37. Werte und Beschriftungen, die für Variablenkategorien angezeigt werden

Die Kategoriewerte zeigen eindeutig an, dass ein Mittelwert von 2,34 zwischen *Mäßig* und *Fast kein* liegt. Durch Anzeigen der Kategoriewerte in der Tabelle wird die Interpretation der angepassten Auswertungsstatistik für Gesamtsummen (z. B. des Mittelwerts) erleichtert.

Diese Anzeigeeinstellung ist eine globale Einstellung, die sich auf die ausgegebenen Pivot-Tabellen aller Prozeduren auswirkt. Die Einstellung wird in allen künftigen Sitzungen beibehalten, bis Sie sie ausdrücklich ändern. So setzen Sie die Einstellung zurück auf die ausschließliche Anzeige von Wertbeschriftungen:

6. Wählen Sie in den Menüs Folgendes aus:
Bearbeiten > Optionen...
7. Klicken Sie im Dialogfeld "Optionen" auf die Registerkarte **Beschriftung der Ausgabe**.
8. Wählen Sie im Gruppenfeld "Beschriftung für Pivot-Tabellen" aus der Dropdown-Liste **Variablenwerte in Beschriftungen anzeigen als** die Option **Beschriftungen** aus.
9. Klicken Sie auf **OK**, um diese Einstellung zu speichern.

Kapitel 8. Auswerten von metrischen Variablen

Auswerten von metrischen Variablen

Es stehen eine Reihe von Auswertungsstatistiken für metrische Variablen zur Verfügung. Neben Häufigkeiten und Prozentsätzen, die für kategoriale Variablen zur Verfügung stehen, sind die folgenden Auswertungsstatistiken für metrische Variablen verfügbar:

- Mittelwert
- Median
- Perzentile
- Summe
- Standardabweichung
- Spannweite
- Minimal- und Maximalwerte

Weitere Informationen finden Sie in „Auswertungsstatistiken für metrische Variablen und angepasste Gesamtsummen für kategoriale Variablen“ auf Seite 9.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Gestapelte metrische Variablen

Durch Stapeln können mehrere metrische Variablen in derselben Tabelle gleichzeitig ausgewertet werden.

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

2. Klicken Sie im Tabellenerstellungsprogramm in der Variablenliste auf *Alter der befragten Person*, klicken Sie anschließend mit gedrückter Steuertaste auf *Letztes abgeschlossenes Schuljahr* und klicken Sie dann mit gedrückter Steuertaste auf *Täglicher Fernsehkonsum in Stunden*, um alle drei Variablen auszuwählen.

3. Ziehen Sie die drei ausgewählten Variablen in den Erstellungsbereich und legen Sie sie im Zeilenbereich ab.

Die drei Variablen werden in der Zeilendimension gestapelt. Da es sich bei allen drei Variablen um metrische Variablen handelt, werden keine Kategorien angezeigt. Die Standardauswertungsstatistik ist der Mittelwert.

4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

	Mean
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

Abbildung 38. Tabelle mit Mittelwerten von gestapelten Variablen

Mehrere Auswertungsstatistiken

In der Standardeinstellung wird für metrische Variablen der Mittelwert angezeigt. Es können aber auch andere Statistiken für metrische Variablen ausgewählt werden und es können mehrere Auswertungsstatistiken gleichzeitig angezeigt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf eine der drei metrischen Variablen und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
3. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option *Median* aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen. (Mit der Pfeilschaltfläche können Sie ausgewählte Statistiken aus der Liste "Statistik" in die Liste "Anzeigen" übertragen. Sie können die Statistiken auch aus der Liste "Statistik" in die Liste "Anzeigen" ziehen und dort ablegen.)
4. Klicken Sie in der Liste "Anzeigen" auf die Zelle **Format** für den Median und wählen Sie aus der Dropdown-Liste der Formate den Eintrag **nnnn** aus.
5. Geben Sie in der Zelle "Dezimalstellen" den Wert 1 ein.
6. Nehmen Sie dieselben Änderungen für den Mittelwert in der Liste "Anzeigen" vor.
7. Klicken Sie auf **Allen zuweisen**, um diese Änderung allen drei metrischen Variablen zuzuweisen.
8. Klicken Sie im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

	Mean	Median
Age of respondent	45.6	42.0
Highest year of school completed	13.3	13.0
Hours per day watching TV	2.9	2.0

Abbildung 39. In der Tabelle mit gestapelten metrischen Variablen werden der Mittelwert und der Median angezeigt.

Anzahl, gültige N und fehlende Werte

Es ist häufig sinnvoll, die bei der Berechnung von Auswertungsstatistiken wie dem Mittelwert verwendete Anzahl der Fälle anzuzeigen. Vermutlich gehen Sie (nicht ganz ohne Grund) davon aus, dass die Auswertungsstatistik *Anzahl* diese Informationen bereitstellt. Wenn Werte fehlen, ergibt dies jedoch keine korrekte Fallbasis. Um eine korrekte Fallbasis zu erhalten, verwenden Sie *Gültige N*.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf eine der drei metrischen Variablen und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
3. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Anzahl** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
4. Wählen Sie anschließend in der Liste "Statistik" die Option **Gültige N** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.

5. Klicken Sie auf **Allen zuweisen**, um diese Änderung allen drei metrischen Variablen zuzuweisen.
6. Klicken Sie im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

	Mean	Median	Count	Valid N
Age of respondent	45.6	42.0	2832	2828
Highest year of school completed	13.3	13.0	2832	2820
Hours per day watching TV	2.9	2.0	2832	2337

Abbildung 40. Anzahl und gültige N

Anzahl liefert für alle drei Variablen denselben Wert: 2.832. Es ist kein Zufall, dass dieser Wert mit der Gesamtanzahl der Fälle in der Datendatei übereinstimmt. Da die metrischen Variablen in keiner kategorialen Variablen verschachtelt sind, stellt *Anzahl* einfach die Gesamtanzahl der Fälle in der Datendatei dar.

Gültige N ergibt dagegen für jede Variable einen anderen Wert, der sich bei *Täglicher Fernsehkonsum in Stunden* erheblich von *Anzahl* unterscheidet. Dieser Unterschied resultiert aus der großen Anzahl von **fehlenden Werten** für diese Variable. Diese sind auf Fälle zurückzuführen, für die in dieser Variablen kein Wert aufgezeichnet wurde oder für die Werte definiert wurden, die fehlende Daten darstellen (z. B. der Code "99", der für die Schwangerschaft bei Männern *nicht zutreffend* angibt).

7. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
8. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf eine der drei metrischen Variablen und wählen Sie im Pop-up-Menü die Option **Auswertungsstatistik** aus.
9. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Anzeigen" die Option **Gültige N** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Statistik" zu übertragen und aus der Liste "Anzeigen" zu entfernen.
10. Wählen Sie in der Liste "Anzeigen" die Option **Anzahl** aus und klicken Sie auf die Pfeilschaltfläche, um sie zurück in die Liste "Statistik" zu verschieben und aus der Liste "Anzeigen" zu entfernen.
11. Wählen Sie in der Liste "Statistik" die Option **Fehlend** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
12. Klicken Sie auf **Allen zuweisen**, um diese Änderung allen drei metrischen Variablen zuzuweisen.
13. Klicken Sie im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

	Mean	Median	Missing
Age of respondent	45.6	42.0	4
Highest year of school completed	13.3	13.0	12
Hours per day watching TV	2.9	2.0	495

Abbildung 41. In der Tabelle mit der Auswertungsstatistik für metrische Variablen wird die Anzahl der fehlenden Werte angezeigt.

In der Tabelle wird nun für jede metrische Variable die Anzahl der fehlenden Werte angezeigt. Daraus wird leicht ersichtlich, dass bei *Täglicher Fernsehkonsum in Stunden* eine große Anzahl von fehlenden Werten vorhanden ist, während bei den anderen beiden Variablen nur sehr wenige zu verzeichnen sind. Dies sollte bei der Beurteilung der Vertrauenswürdigkeit dieser Auswertung für die Variable berücksichtigt werden.

Unterschiedliche Auswertungen für unterschiedliche Variablen

In einer gestapelten Tabelle können nicht nur mehrere Auswertungsstatistiken, sondern auch unterschiedliche Auswertungsstatistiken für unterschiedliche metrische Variablen angezeigt werden. Aus der vorangegangenen Tabelle geht beispielsweise hervor, dass nur eine der drei Variablen über eine große Anzahl fehlender Werte verfügt. Möglicherweise empfiehlt es sich daher, nur für diese Variable die Anzahl der fehlenden Werte in die Tabelle aufzunehmen.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie in der Tabellenvorschau im Erstellungsbereich auf *Alter der befragten Person* und klicken Sie anschließend mit gedrückter Steuertaste auf *Letztes abgeschlossenes Schuljahr*, um beide Variablen auszuwählen.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf eine der beiden Variablen und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
4. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Anzeigen" die Option **Fehlend** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Statistik" zu übertragen und aus der Liste "Anzeigen" zu entfernen.
5. Klicken Sie auf **Der Auswahl zuweisen**, um die Änderung ausschließlich den beiden ausgewählten Variablen zuzuweisen.

Die Platzhalter in den Datenzellen der Tabelle zeigen an, dass die Anzahl der fehlenden Werte nur für *Täglicher Fernsehkonsum in Stunden* angezeigt wird.

6. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

	Mean	Median	Missing
Age of respondent	45.6	42.0	
Highest year of school completed	13.3	13.0	
Hours per day watching TV	2.9	2.0	495

Abbildung 42. Tabelle mit unterschiedlichen Auswertungsstatistiken für unterschiedliche Variablen

Diese Tabelle liefert zwar die gewünschten Informationen, das Layout erschwert jedoch die Interpretation der Tabelle. Beim Lesen der Tabelle könnte den Eindruck entstehen, dass die leeren Zellen in der Spalte *Fehlend* bedeuten, dass 0 fehlende Werte für die entsprechenden Variablen vorliegen.

7. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
8. Wählen Sie im Tabellenerstellungsprogramm im Gruppenfeld "Auswertungsstatistik" den Eintrag **Zeilen** aus der Dropdown-Liste "Position" aus.
9. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

Age of respondent	Mean	45.6
	Median	42.0
Highest year of school completed	Mean	13.3
	Median	13.0
Hours per day watching TV	Mean	2.9
	Median	2.0
	Missing	495

Abbildung 43. Die Auswertungsstatistiken und Variablen werden in der Zeilendimension angezeigt.

Jetzt geht aus der Tabelle eindeutig hervor, dass nur für eine Variable die Anzahl der fehlenden Werte aufgeführt wird.

Gruppierte Auswertungen in Kategorien

Kategoriale Variablen können als Gruppierungsvariablen verwendet werden, um Auswertungen für metrische Variablen in den durch die Kategorien der kategorialen Variablen definierten Gruppen anzuzeigen.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste und legen Sie sie im Spaltenbereich des Erstellungsbereichs ab.

Wenn Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf *Geschlecht* klicken, werden Sie feststellen, dass der Befehl **Auswertungsstatistik** im Popup-Menü inaktiviert ist. Die Ursache hierfür ist, dass in einer Tabelle mit metrischen Variablen diese immer die Quellenvariablen für die Statistiken darstellen.

3. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Gender	
		Male	Female
Age of respondent	Mean	44.6	46.3
	Median	42.0	43.0
Highest year of school completed	Mean	13.4	13.2
	Median	13.0	13.0
Hours per day watching TV	Mean	2.8	2.9
	Median	2.0	2.0
	Missing	213	282

Abbildung 44. Gruppierte Auswertungen von metrischen Variablen anhand einer kategorialen Spaltenvariablen

Anhand dieser Tabelle lassen sich die Durchschnittswerte (Mittelwert und Median) für Frauen und Männer leicht vergleichen. Sie zeigt eindeutig, dass die Unterschiede relativ gering sind. Dies ist zwar möglicherweise nicht sehr interessant, kann aber eine nützliche Information darstellen.

Mehrere Gruppierungsvariablen

Die Gruppen können durch Verschachtelung und/oder gleichzeitiger Verwendung von kategorialen Zeilen- und Spaltenvariablen als Gruppierungsvariablen weiter unterteilt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie *Nachrichten aus dem Internet* aus der Variablenliste und legen Sie sie im äußersten linken Rand des Zeilenbereichs des Erstellungsbereichs ab. Stellen Sie sicher, dass Sie die Variable so ablegen, dass wirklich alle drei metrischen Variablen darin verschachtelt sind, nicht nur eine der Variablen.

Es mag Situationen geben, in denen eine Verschachtelung wie im zweiten Beispiel oben erwünscht ist. Im vorliegenden Beispiel trifft dies jedoch nicht zu.

3. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

				Gender	
				Male	Female
Get news from internet	No	Age of respondent	Mean	47.0	48.8
			Median	45.0	46.0
		Highest year of school completed	Mean	13.4	13.1
			Median	13.0	12.0
		Hours per day watching TV	Mean	3.2	3.4
			Median	2.0	3.0
	Yes	Age of respondent	Missing	213	282
			Mean	38.7	41.1
			Median	35.0	38.0
		Highest year of school completed	Mean	13.2	13.3
			Median	13.0	13.0
		Hours per day watching TV	Mean	2.1	2.1
			Median	2.0	2.0
		Missing	0	0	

Abbildung 45. Auswertungen von metrischen Variablen gruppiert nach kategorialen Zeilen- und Spaltenvariablen

Verschachteln von kategorialen in metrischen Variablen

Die vorangegangene Tabelle liefert zwar die gewünschten Informationen, bietet sie jedoch in einem Format an, das die Interpretation nicht gerade erleichtert. So kann beispielsweise anhand der Tabelle das durchschnittliche Alter der Personen verglichen werden, die Nachrichten aus dem Internet bzw. nicht aus dem Internet erfahren. Der Vergleich ließe sich jedoch leichter durchführen, wenn die Werte nebeneinander angezeigt würden. Durch Vertauschen der Positionen der beiden Zeilenvariablen und Verschachteln der kategorialen Gruppierungsvariablen in den drei metrischen Variablen kann die Tabellendarstellung verbessert werden. Bei metrischen Variablen hat die Verschachtelungsebene keinen Einfluss auf die Quellenvariable der Statistik. Die metrische Variable ist unabhängig von der Verschachtelungsebene immer die Quellenvariable für die Statistik.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie in der Tabellenvorschau im Erstellungsbereich auf *Alter der befragten Person*, klicken Sie anschließend mit gedrückter Steuertaste auf *Letztes abgeschlossenes Schuljahr* und klicken Sie dann mit gedrückter Steuertaste auf *Täglicher Fernsehkonsum in Stunden*, um alle drei metrischen Variablen auszuwählen.
3. Ziehen Sie die drei metrischen Variablen und legen Sie sie so am äußersten linken Rand des Zeilenbereichs ab, dass die kategoriale Variable *Nachrichten aus dem Internet* in jeder der drei metrischen Variablen verschachtelt ist.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

				Gender	
				Male	Female
Age of respondent	Get news from internet	No	Mean	47.0	48.8
			Median	45.0	46.0
		Yes	Mean	38.7	41.1
			Median	35.0	38.0
Highest year of school completed	Get news from internet	No	Mean	13.4	13.1
			Median	13.0	12.0
		Yes	Mean	13.2	13.3
			Median	13.0	13.0
Hours per day watching TV	Get news from internet	No	Mean	3.2	3.4
			Median	2.0	3.0
			Missing	213	282
		Yes	Mean	2.1	2.1
			Median	2.0	2.0
			Missing	0	0

Abbildung 46. In gestapelten metrischen Variablen verschachtelte kategoriale Zeilenvariable

Die Wahl der Verschachtelungsreihenfolge hängt von den Beziehungen oder Vergleichen ab, die in der Tabelle hervorgehoben werden sollen. Ein Ändern der Verschachtelungsreihenfolge der metrischen Variablen zieht keine Änderung der Werte in den Auswertungsstatistiken nach sich. Lediglich die relative Position in der Tabelle wird geändert.

Kapitel 9. Konfidenzintervalle

Konfidenzintervalle und Standardfehler sind für viele Tabellenstatistiken verfügbar.

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Verschieben Sie *Höchster Schulabschluss* im Tabellenerstellungsprogramm in den Bereich für die Zeilen im Erstellungsbereich.
3. Klicken Sie auf **Auswertungsstatistik...**
4. Klicken Sie in der Liste **Statistik** im Dialogfeld **Auswertungsstatistik** auf das Symbol neben **Anzahl**, um die Liste der anzahlbezogenen Statistiken zu erweitern.
5. Verschieben Sie die folgenden Statistiken in den Bereich **Anzeige**: Anzahl als Spalten (%), Untere KG für Anzahl als Spalten (%), Obere KG für Anzahl als Spalten (%) und Standardfehler der Anzahl als Spalten (%).
6. Geben Sie in der Gruppe **Konfidenzintervalle** für **Stufe (%)** den Wert 99 ein.
7. Klicken Sie auf **Der Auswahl zuweisen** und klicken Sie dann auf **Schließen**.
8. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	99.0% Lower CL for Column N %	99.0% Upper CL for Column N %	Standard Error of Column N %
Highest degree	LT High school	430	15.2%	13.6%	17.0%	0.7%
	High school	1500	53.2%	50.7%	55.6%	0.9%
	Junior college	209	7.4%	6.2%	8.7%	0.5%
	Bachelor	478	16.9%	15.2%	18.8%	0.7%
	Graduate	205	7.3%	6.1%	8.6%	0.5%

Abbildung 47. Häufigkeitstabelle, Spaltenprozentsätze und Konfidenzintervalle

9. Rufen Sie das Dialogfeld **Benutzerdefinierte Tabellen** erneut auf und klicken Sie auf **Auswertungsstatistik...**
10. Ändern Sie für die untere Konfidenzgrenze die Beschriftung in "Untere Konfidenzgrenze (&[Konfidenzniveau])". Die Zeichenfolge "&[Konfidenzniveau]" fügt den Wert des angegebenen Konfidenzniveaus an dieser Position in die Beschriftung ein.
11. Klicken Sie auf **Der Auswahl zuweisen** und klicken Sie dann auf **Schließen**.
12. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	Lower Confidence Limit (99.0%)	99.0% Upper CL for Column N %	Standard Error of Column N %
Highest degree	LT High school	430	15.2%	13.6%	17.0%	0.7%
	High school	1500	53.2%	50.7%	55.6%	0.9%
	Junior college	209	7.4%	6.2%	8.7%	0.5%
	Bachelor	478	16.9%	15.2%	18.8%	0.7%
	Graduate	205	7.3%	6.1%	8.6%	0.5%

Abbildung 48. Tabelle mit geänderter Beschriftung für das Konfidenzintervall

Kapitel 10. Teststatistiken

Teststatistiken

Es stehen drei verschiedene Tests auf Signifikanz zum Untersuchen der Beziehung zwischen Zeilen- und Spaltenvariablen zur Verfügung. In diesem Kapitel wird die Ausgabe dieser Tests erläutert, wobei besonderes Augenmerk auf die Auswirkungen von verschachtelten und gestapelten Variablen gerichtet wird. Weitere Informationen finden Sie in Kapitel 3, „Stapeln, Verschachteln und Schichten von kategorialen Variablen“, auf Seite 31.

Beispieldatendatei

Für die Beispiele dieses Kapitels wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

Tests auf Unabhängigkeit (Chi-Quadrat)

Mit dem Chi-Quadrat-Test auf Unabhängigkeit kann ermittelt werden, ob eine Beziehung zwischen zwei kategorialen Variablen vorliegt. So kann beispielsweise ermittelt werden, ob es einen Zusammenhang zwischen *Beschäftigungsstatus* und *Familienstand* gibt.

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Ziehen Sie die Variable *Beschäftigungsstatus* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich für die Zeilen im Erstellungsbereich.
3. Ziehen Sie die Variable *Familienstand* aus der Variablenliste und legen Sie sie im Spaltenbereich ab.
4. Wählen Sie **Zeilen** als Position für die Auswertungsstatistiken.
5. Wählen Sie *Beschäftigungsstatus* aus und klicken Sie im Gruppenfeld "Definieren" auf **Auswertungsstatistik**.
6. Wählen Sie in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und übertragen Sie sie in die Liste "Anzeigen".
7. Klicken Sie auf **Der Auswahl zuweisen**.
8. Klicken Sie im Dialogfeld "Benutzerdefinierte Tabellen" auf die Registerkarte **Teststatistiken**.
9. Wählen Sie **Tests auf Unabhängigkeit (Chi-Quadrat)** aus.
10. Klicken Sie auf **OK**, um die Tabelle zu erstellen und den Chi-Quadrat-Test durchzuführen.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
Labor force status	Working full time	Count 778	44	295	58	392
		Column % 57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	Count 138	20	35	9	102
		Column % 10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	Count 23	2	9	1	11
		Column % 1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	Count 13	3	10	0	32
		Column % 1.0%	1.1%	2.2%	.0%	4.8%
	Retired	Count 168	150	53	6	17
		Column % 12.5%	53.0%	11.9%	6.5%	2.6%
	School	Count 9	1	7	2	60
		Column % .7%	.4%	1.6%	2.2%	9.0%
	Keeping house	Count 200	55	25	13	35
		Column % 14.9%	19.4%	5.6%	14.0%	5.3%
	Other	Count 16	8	12	4	14
		Column % 1.2%	2.8%	2.7%	4.3%	2.1%

Abbildung 49. Beschäftigungsstatus nach Familienstand

Bei dieser Tabelle handelt es sich um eine Kreuztabelle von *Beschäftigungsstatus* nach *Familienstand*, in der die Häufigkeiten und Spaltenanteile als Auswertungsstatistiken angezeigt werden. Die Spaltenanteile werden so berechnet, dass die Summe jeder Spalte 100 % ergibt. Wenn zwischen den beiden Variablen kein Zusammenhang besteht, sollten sich in jeder Zeile ähnliche Anteile in allen Spalten befinden. Bei den Anteilen scheint es Differenzen zu geben. Sie können jedoch den Chi-Quadrat-Test überprüfen, um sich zu vergewissern.

		Marital status
Labor force status	Chi-square	729.242
	df	28
	Sig.	.000*

*. The Chi-square statistic is significant at the 0.05 level.

Abbildung 50. Chi-Quadrat-Test nach Pearson

Beim Test auf Unabhängigkeit wird davon ausgegangen, dass zwischen *Beschäftigungsstatus* und *Familienstand* keine Beziehung besteht, also die Spaltenanteile in allen Spalten gleich sind und beobachtete Abweichungen auf zufällige Variation zurückzuführen sind. Bei der Chi-Quadrat-Statistik wird die Gesamtabweichung zwischen den beobachteten Zellenhäufigkeiten und den bei Gleichheit der Spaltenanteile über alle Spalten erwarteten Häufigkeiten gemessen. Eine größere Chi-Quadrat-Statistik zeigt eine größere Abweichung zwischen den beobachteten und den erwarteten Zellenhäufigkeiten an. Dies deutet darauf hin, dass die Spaltenanteile nicht gleich sind, dass die Hypothese der Unabhängigkeit somit falsch ist und dass deshalb ein Zusammenhang zwischen *Beschäftigungsstatus* und *Familienstand* besteht.

Die berechnete Chi-Quadrat-Statistik weist den Wert 729,242 auf. Um festzustellen, ob dies ausreicht, um die Hypothese der Unabhängigkeit zu widerlegen, wird der Signifikanzwert der Statistik berechnet. Der Signifikanzwert ist die Wahrscheinlichkeit, dass eine Zufallsvariable aus einer Chi-Quadrat-Verteilung mit 28 Freiheitsgraden größer als 729,242 ist. Da dieser Wert kleiner als das auf der Registerkarte "Teststatistiken" angegebene Alpha-Niveau ist, kann die Hypothese der Unabhängigkeit auf dem Niveau 0,05 widerlegt werden. Somit besteht tatsächlich ein Zusammenhang zwischen *Beschäftigungsstatus* und *Familienstand*.

Auswirkungen von verschachtelten und gestapelten Variablen auf Tests auf Unabhängigkeit

Die Regel für Tests auf Unabhängigkeit besagt, dass für jede innerste Untertabelle ein separater Test durchgeführt wird. Um die Auswirkung von verschachtelten Variablen auf die Tests zu verdeutlichen, betrachten Sie das vorhergehende Beispiel, wobei jedoch *Familienstand* in Ebenen von *Geschlecht* verschachtelt wird.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste in den Erstellungsbereich und legen Sie sie im Spaltenbereich über *Familienstand* ab.
3. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{*.1.2}	.000 ^{*.1.2}

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Abbildung 51. Chi-Quadrat-Test nach Pearson

Wenn *Familienstand* in Ebenen von *Geschlecht* verschachtelt wird, werden zwei Tests durchgeführt: einer für jede Ebene von *Geschlecht*. Der Signifikanzwert für die einzelnen Tests zeigt an, dass die Hypothese der Unabhängigkeit zwischen *Familienstand* und *Beschäftigungsstatus* sowohl für Männer als auch für Frauen widerlegt werden kann. Den Anmerkungen der Tabelle kann jedoch entnommen werden, dass über 20 % der Zellen jeder Tabelle erwartete Häufigkeiten von weniger als 5 aufweisen und die niedrigste erwartete Zellenhäufigkeit unter 1 liegt. Die Anmerkungen weisen darauf hin, dass die Annahmen des Chi-Quadrat-Tests möglicherweise nicht auf diese Tabellen zutreffen und die Ergebnisse der Tests somit nicht zuverlässig sind.

Hinweis: Die Fußnoten sind möglicherweise aufgrund der Zellenbegrenzungen nicht sichtbar. Sie können die Fußnoten anzeigen, indem Sie die Ausrichtung der Zellen im Dialogfeld "Zelleneigenschaften" ändern.

So verdeutlichen Sie die Auswirkung von gestapelten Variablen auf die Tests:

4. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
5. Ziehen Sie die Variable *Höchster Schulabschluss* aus der Variablenliste und legen Sie sie im Zeilenbereich unter *Beschäftigungsstatus* ab.
6. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{a,1,2}	.000 ^{a,1,2}
Highest degree	Chi-square	43.844	105.506
	df	16	16
	Sig.	.000 ^a	.000 ^a

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.

2. The minimum expected cell count in this sub-table is less than one.

Abbildung 52. Chi-Quadrat-Test nach Pearson

Wenn *Höchster Schulabschluss* mit *Beschäftigungsstatus* gestapelt wird, werden vier Tests durchgeführt: für jede Ebene von *Geschlecht* ein Test auf Unabhängigkeit von *Familienstand* und *Beschäftigungsstatus* und ein Test auf Unabhängigkeit von *Familienstand* und *Höchster Schulabschluss*. Die Testergebnisse für *Familienstand* und *Beschäftigungsstatus* sind mit den vorherigen Testergebnissen identisch. Die Testergebnisse für *Familienstand* und *Höchster Schulabschluss* weisen darauf hin, dass diese Variablen nicht unabhängig sind.

Vergleichen der Spaltenmittel

Mit den Spaltenmitteltests wird ermittelt, ob eine Beziehung zwischen einer kategorialen Variablen in den Spalten und einer stetigen Variablen in den Zeilen vorliegt. Außerdem können Sie mit den Testergebnissen die relative Ordnung der Kategorien der kategorialen Variablen hinsichtlich des Mittelwerts der stetigen Variablen ermitteln. So können Sie beispielsweise ermitteln, ob zwischen *Täglicher Fernsehkonsum in Stunden* und *Nachrichten aus Zeitungen* ein Zusammenhang besteht.

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Klicken Sie auf **Zurücksetzen**, um die Standardeinstellungen aller Registerkarten wiederherzustellen.
3. Ziehen Sie im Tabellenerstellungsprogramm die Variable *Täglicher Fernsehkonsum in Stunden* aus der Variablenliste und legen Sie sie im Zeilenbereich des Erstellungsbereichs ab.
4. Ziehen Sie die Variable *Nachrichten aus Zeitungen* aus der Variablenliste und legen Sie sie im Spaltenbereich ab.
5. Wählen Sie *Täglicher Fernsehkonsum in Stunden* aus und klicken Sie im Gruppenfeld "Definieren" auf **Auswertungsstatistik**.
6. Wählen Sie **nnnn** als Format aus.
7. Wählen Sie **2** als Anzahl der angezeigten Dezimalstellen aus. Beachten Sie, dass das Format nun als **nnnn,nn** dargestellt wird.
8. Klicken Sie auf **Der Auswahl zuweisen**.
9. Klicken Sie im Dialogfeld "Benutzerdefinierte Tabellen" auf die Registerkarte **Teststatistiken**.
10. Wählen Sie **Spaltenmittel vergleichen (T-Tests)** aus.
11. Klicken Sie auf **OK**, um die Tabelle zu erstellen und die Spaltenmitteltests durchzuführen.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92	2.74

Abbildung 53. Nachrichten aus Zeitungen nach täglichem Fernsehkonsum in Stunden

Diese Tabelle zeigt den Mittelwert von *Täglicher Fernsehkonsum in Stunden* für Personen, die Nachrichten aus Zeitungen bzw. nicht aus Zeitungen erfahren. Die beobachtete Differenz in diesen Mittelwerten deutet darauf hin, dass Personen, die Nachrichten nicht aus Zeitungen erfahren, zirka 0,18 Stunden länger fernsehen als Personen, die Nachrichten aus Zeitungen erfahren. Um herauszufinden, ob diese Differenz auf zufällige Variation zurückzuführen ist, überprüfen Sie die Spaltenmitteltests.

	Get news from newspapers	
	No	Yes
	(A)	(B)
Hours per day watching TV		

Abbildung 54. Vergleich der Spaltenmittel

In der Tabelle für den Spaltenmitteltest wird jeder Kategorie der Spaltenvariablen ein Buchstabenschlüssel zugewiesen. Bei *Nachrichten aus Zeitungen* wird der Kategorie *Nein* der Buchstabe A zugewiesen und *Ja* erhält den Buchstaben B. Für jedes Spaltenpaar werden die Spaltenmittel mithilfe des *T*-Tests verglichen. Da nur zwei Spalten vorliegen, wird nur ein Test durchgeführt. Für jedes signifikante Paar wird der Schlüssel der Kategorie mit dem kleineren Mittelwert unterhalb der Kategorie mit dem größeren Mittelwert ausgegeben. Da in den Zellen der Tabelle keine Schlüssel ausgegeben werden, bedeutet dies, dass die Spaltenmittel statistisch gesehen nicht differieren.

Signifikanzergebnisse durch Notation im APA-Stil

Wenn Sie nicht möchten, dass die Signifikanzergebnisse in einer separaten Tabelle angezeigt werden, können Sie diese auch in der Haupttabelle anzeigen. Signifikanzergebnisse werden anhand einer Notation im APA-Stil mit tiefgestellten Buchstaben identifiziert. Führen Sie die obigen Schritte zum Vergleich von Spaltenmitteln aus, nehmen Sie jedoch auf der Registerkarte "Teststatistiken" folgende Änderung vor:

1. Wählen Sie im Bereich "Signifikante Unterschiede identifizieren" die Option **In der Haupttabelle unter Verwendung von Teilscripts im APA-Stil** aus.
2. Klicken Sie auf **OK**, um die Tabelle zu erstellen und die Spaltenmitteltests anhand von Notation im APA-Stil durchzuführen.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92 _a	2.74 _a

Abbildung 55. Vergleiche von Spaltenmitteln anhand von Notation im APA-Stil

In der Tabelle für den Spaltenmitteltest wird den Kategorien der Spaltenvariablen ein tiefgestellter Buchstabe zugewiesen. Die Spaltenmittel werden für jedes Spaltenpaar mithilfe eines *T*-Tests verglichen. Wenn sich ein Wertepaar stark unterscheidet, werden den Werten *unterschiedliche* tiefgestellte Buchstaben zugewiesen. Da nur zwei Spalten vorliegen, wird nur ein Test durchgeführt. Da die Spaltenmittel in diesem Beispiel über denselben tiefgestellten Buchstaben verfügen, unterscheiden sich die Spaltenmittel statistisch gesehen nicht.

Auswirkungen von verschachtelten und gestapelten Variablen auf Spaltenmitteltests

Die Regel für Spaltenmitteltests besagt, dass für jede innerste Untertabelle eine separate Reihe paarweiser Tests durchgeführt wird. Um die Auswirkung von verschachtelten Variablen auf die Tests zu verdeutlichen, betrachten Sie das vorhergehende Beispiel, wobei jedoch *Täglicher Fernsehkonsum in Stunden* in Ebenen von *Beschäftigungsstatus* verschachtelt wird.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Beschäftigungsstatus* aus der Variablenliste und legen Sie sie im Zeilenbereich im Erstellungsbereich ab.
3. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

			Get news from newspapers	
			No	Yes
			(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B	
	Working part-time	Hours per day watching TV		
	Temporarily not working	Hours per day watching TV		
	Unemployed, laid off	Hours per day watching TV		
	Retired	Hours per day watching TV		
	School	Hours per day watching TV		
	Keeping house	Hours per day watching TV		
	Other	Hours per day watching TV		

Abbildung 56. Vergleich der Spaltenmittel

Wenn *Täglicher Fernsehkonsum in Stunden* in Ebenen von *Beschäftigungsstatus* verschachtelt ist, werden sieben Reihen von Spaltenmitteltests durchgeführt: eine für jede Ebene von *Beschäftigungsstatus*. Den Kategorien von *Nachrichten aus Zeitungen* werden dieselben Buchstabenschlüssel zugewiesen. Für Befragte mit Beschäftigung in *Vollzeit* wird der Schlüssel B in der Spalte A angezeigt. Das bedeutet, dass bei Vollzeitbeschäftigten der Mittelwert von *Täglicher Fernsehkonsum in Stunden* bei den Personen niedriger ist, die *Nachrichten aus Zeitungen* erfahren. Da in den Spalten keine anderen Schlüssel angezeigt werden, können Sie daraus schließen, dass bei den Spaltenmitteln keine weiteren statistisch signifikanten Differenzen vorliegen.

Bonferroni-Korrekturen. Wenn mehrere Tests durchgeführt werden, wird die Bonferroni-Korrektur auf die Spaltenmitteltests angewendet, um sicherzustellen, dass das auf der Registerkarte "Teststatistiken" angegebene Alpha-Niveau (Rate der fälschlicherweise positiven Ergebnisse) für jede *Testreihe* gilt. In dieser Tabelle wurden keine Bonferroni-Korrekturen angewendet, denn obwohl sieben Testreihen durchgeführt wurden, wurde in jeder Reihe nur ein Spaltenpaar verglichen.

So verdeutlichen Sie die Auswirkung von gestapelten Variablen auf die Tests:

4. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
5. Ziehen Sie die Variable *Nachrichten aus dem Internet* aus der Variablenliste und legen Sie sie im Spaltenbereich links neben *Nachrichten aus Zeitungen* ab.
6. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

			Get news from internet		Get news from newspapers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B		B	
	Working part-time	Hours per day watching TV	B			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	B			
	Retired	Hours per day watching TV	B			
	School	Hours per day watching TV	B			
	Keeping house	Hours per day watching TV	B			
	Other	Hours per day watching TV	B			

Abbildung 57. Vergleich der Spaltenmittel

Wenn Nachrichten aus dem Internet mit Nachrichten aus Zeitungen gestapelt ist, werden 14 Reihen von Spaltenmitteltests durchgeführt: eine für jede Ebene von *Beschäftigungsstatus* sowohl für Nachrichten aus dem Internet als auch für Nachrichten aus Zeitungen. Wiederum werden keine Bonferroni-Korrekturen vorgenommen, da in jeder Reihe nur ein Spaltenpaar verglichen wird. Die Tests für Nachrichten aus Zeitungen sind dieselben wie vorher. Bei Nachrichten aus dem Internet wird der Kategorie *Nein* der Buchstabe A zugewiesen und *Ja* erhält den Buchstaben B. Außer für die Befragten, die derzeit "zeitweise unbeschäftigt" sind, wird in jeder Reihe von Spaltenmitteltests der Schlüssel B in der Spalte A ausgegeben. Dies bedeutet, dass der Mittelwert von *Täglicher Fernsehkonsum in Stunden* bei Personen, die Nachrichten aus dem Internet erfahren, niedriger ist als bei Personen, die Nachrichten nicht aus der Zeitung erfahren. Für die Testreihe für *Zeitweise unbeschäftigt* werden keine Schlüssel ausgegeben; somit differieren die Spaltenmittel für diese Befragten statistisch gesehen nicht.

Vergleichen der Spaltenanteile

Mit den Spaltenanteiltests wird die relative Ordnung der Kategorien der kategorialen Spaltenvariablen hinsichtlich der Kategorieanteile der kategorialen Zeilenvariablen ermittelt. Nachdem Sie beispielsweise mit einem Chi-Quadrat-Test festgestellt haben, dass *Beschäftigungsstatus* und *Familienstand* nicht unabhängig sind, können Sie ermitteln, welche Zeilen und Spalten für diese Beziehung verantwortlich sind.

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Klicken Sie auf **Zurücksetzen**, um die Standardeinstellungen aller Registerkarten wiederherzustellen.
3. Ziehen Sie die Variable *Beschäftigungsstatus* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich für die Zeilen im Erstellungsbereich.
4. Ziehen Sie die Variable *Familienstand* aus der Variablenliste und legen Sie sie im Spaltenbereich ab.
5. Wählen Sie *Beschäftigungsstatus* aus und klicken Sie im Gruppenfeld "Definieren" auf **Auswertungstatistik**.
6. Wählen Sie in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und übertragen Sie sie in die Liste "Anzeigen".
7. Inaktivieren Sie in der Liste "Anzeigen" die Option **Anzahl**.
8. Klicken Sie auf **Der Auswahl zuweisen**.

9. Klicken Sie im Dialogfeld "Benutzerdefinierte Tabellen" auf die Registerkarte **Teststatistiken**.
10. Wählen Sie **Spaltenanteile vergleichen (Z-Tests)** aus.
11. Klicken Sie auf **OK**, um die Tabelle zu erstellen und die Spaltenanteiletests durchzuführen.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column %	Column %	Column %	Column %	Column %
Labor force status	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	12.5%	53.0%	11.9%	6.5%	2.6%
	School	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Abbildung 58. Beschäftigungsstatus nach Familienstand

Bei dieser Tabelle handelt es sich um eine Kreuztabelle von *Beschäftigungsstatus* nach *Familienstand*, in der die Spaltenanteile als Auswertungsstatistik angezeigt werden.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Labor force status	Working full time	B		A B	B	B
	Working part-time					A B C
	Temporarily not working					A B
	Unemployed, laid off					A B
	Retired	E	A C D E	E		
	School					A B C
	Keeping house	C E	C E		C E	
	Other					

Abbildung 59. Vergleich der Spaltenanteile

In der Tabelle für den Spaltenanteiletest wird jeder Kategorie der Spaltenvariablen ein Buchstabenschlüssel zugewiesen. Bei *Familienstand* wird der Kategorie *verheiratet* der Buchstabe A zugewiesen, *verwitwet* wird der Buchstabe B zugewiesen usw., bis zur Kategorie *ledig*, der der Buchstabe E zugewiesen wird. Für jedes Spaltenpaar werden die Spaltenanteile mit einem Z-Test verglichen. Es werden sieben Reihen von Spaltenanteiletests durchgeführt, eine für jede Ebene von *Beschäftigungsstatus*. Da bei *Familienstand* fünf Ebenen vorliegen, werden in jeder Testreihe $(5 \cdot 4) / 2 = 10$ Spaltenpaare verglichen und zum Korrigieren der Signifikanzwerte werden Bonferroni-Korrekturen verwendet. Für jedes signifikante Paar wird der Schlüssel der kleineren Kategorie unterhalb der Kategorie mit dem größeren Anteil ausgegeben.

Für die Testreihen mit *Vollzeit* wird der Schlüssel B in jeder der anderen Spalten angezeigt. Außerdem wird der Schlüssel A in der Spalte C angezeigt. In anderen Spalten werden keine weiteren Schlüssel gezeigt. Es kann geschlossen werden, dass der Anteil der geschiedenen Personen, die einer Vollzeitbeschäftigung nachgehen, größer ist als der Anteil der verheirateten Personen, die Vollzeit arbeiten. Dieser Anteil ist wiederum größer als der Anteil der verwitweten Personen, die in Vollzeit tätig sind. Der Anteil der Personen, die getrennt lebend oder ledig sind und Vollzeit arbeiten, kann nicht vom Anteil der Personen unterschieden werden, die geschieden oder verheiratet sind und eine Vollzeitbeschäftigung haben. Diese Anteile sind jedoch größer als der Anteil der vollzeitbeschäftigten verwitweten Personen.

Für die Tests im Zusammenhang mit *Teilzeit* oder *Ausbildung* werden die Schlüssel A, B und C in der Spalte E angezeigt. In anderen Spalten werden keine weiteren Schlüssel gezeigt. Somit sind die Anteile von Personen, die ledig sind und zur Schule gehen oder Teilzeit arbeiten, größer als die Anteile von verheirateten, verwitweten oder geschiedenen Personen, die zur Schule gehen oder teilzeitbeschäftigt sind.

Für die Tests im Zusammenhang mit *Zeitweise unbeschäftigt* oder *Andere* werden in keiner Spalten weitere Schlüssel ausgegeben. Somit gibt es keine erkennbare Differenz bei den Anteilen der verheirateten, verwitweten, geschiedenen, getrennt lebenden oder ledigen Personen, die zeitweise unbeschäftigt sind oder sich in einer anderen nicht kategorisierten Beschäftigungssituation befinden.

Die Tests im Zusammenhang mit *Ruhestand* zeigen, dass der Anteil der verwitweten Personen im Ruhestand größer ist als die Anteile aller anderen Kategorien von Personen im Ruhestand. Außerdem sind die Anteile von verheirateten und geschiedenen Personen im Ruhestand größer als die Anteile von ledigen Personen im Ruhestand.

Es gibt größere Anteile von verheirateten, verwitweten und getrennt lebenden Personen, die den Haushalt führen, als Anteile von geschiedenen oder ledigen Personen, die den Haushalt führen.

Der Anteil der ledigen Personen, die *arbeitslos, entlassen* sind, ist höher als der Anteil der verheirateten bzw. verwitweten arbeitslosen Personen. Beachten Sie auch, dass die Spalte *getrennt lebend* mit einem "." gekennzeichnet ist, was darauf hinweist, dass der beobachtete Anteil von getrennt lebenden Personen in der Zeile *arbeitslos, entlassen* 0 oder 1 beträgt und deshalb für arbeitslose Befragte kein Vergleich mit dieser Spalte erfolgen kann.

Zusammenführen von Signifikanzergebnissen in der Haupttabelle

Wenn Sie nicht möchten, dass die Signifikanzergebnisse in einer separaten Tabelle angezeigt werden, können Sie diese auch in der Haupttabelle anzeigen. Führen Sie die obigen Schritte zum Vergleich von Spaltenanteilen aus, nehmen Sie jedoch auf der Registerkarte "Teststatistiken" folgende Änderung vor:

1. Wählen Sie im Bereich "Signifikante Unterschiede identifizieren" die Option **In der Haupttabelle** aus.
2. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Abbildung 60. In der Haupttabelle zusammengeführte Signifikanzergebnisse

3. Doppelklicken Sie im Fenster des Viewers auf die Tabelle, um diese zu aktivieren.
4. Fahren Sie mit der Maus über eine der Spaltenbeschriftungszellen, die die Buchstabenschlüssel enthalten. Fahren Sie beispielsweise mit der Maus über die Zelle mit der Beschriftung "(a)".

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Abbildung 61. Hervorheben aller Zellen, die den ausgewählten Buchstabenschlüssel enthalten

Alle Zellen, die diesen Schlüssel enthalten, werden in der Tabelle hervorgehoben.

5. Klicken Sie mit der rechten Maustaste auf eine Spaltenbeschriftungszelle mit dem Buchstabenschlüssel. Wählen Sie im Kontextmenü **Auswählen > Alle Zellen mit diesem Signifikanzschlüssel auswählen** aus.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(a)	(b)	(c)	(d)	(e)
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% b	15.5%	66.1% a b	62.4% b	59.1% b
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4% a b c
	Temporarily not working	1.7%	0.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	0.0% ¹	4.8% a b
	Retired	12.5% e	53.0% a c d e	11.9% e	6.5%	2.6%
	School	0.7%	0.4%	1.6%	2.2%	9.0% a b c
	Keeping house	14.9% c e	19.4% c e	5.6%	14.0% c e	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Abbildung 62. Auswählen aller Zellen, die den ausgewählten Buchstabenschlüssel enthalten

Alle Zellen, die diesen Schlüssel enthalten, werden in der Tabelle ausgewählt.

Auswirkungen von verschachtelten und gestapelten Variablen auf Spaltenanteiletests

Die Regel für Spaltenanteiletests besagt, dass für jede innerste Untertabelle eine separate Reihe paarweiser Tests durchgeführt wird. Um die Auswirkung von verschachtelten Variablen auf die Tests zu verdeutlichen, betrachten Sie das vorhergehende Beispiel, wobei jedoch *Beschäftigungsstatus* in Ebenen von *Geschlecht* verschachtelt wird.

1. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste und legen Sie sie im Spaltenbereich im Erstellungsbereich ab.
3. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
			Keeping house	C E	C E		C	
		Other						

Abbildung 63. Vergleich der Spaltenanteile

Wenn *Beschäftigungsstatus* in Ebenen von *Geschlecht* verschachtelt ist, werden 14 Reihen von Spaltenanteiletests durchgeführt: eine für jede Ebene von *Beschäftigungsstatus* für jede Ebene von *Geschlecht*. Den Kategorien von *Familienstand* werden dieselben Buchstabenschlüssel zugewiesen.

Beachten Sie die folgenden Punkte bei den Tabellenergebnissen:

- Wenn mehr Tests durchgeführt werden, sind mehr Spalten mit einem Spaltenanteil von null vorhanden. Diese kommen bei getrennt lebenden Befragten und verwitweten Männern am häufigsten vor.
- Die Spaltendifferenzen bei den Befragten, die den *Haushalt* führen, können scheinbar gänzlich auf Frauen zurückgeführt werden.

So verdeutlichen Sie die Auswirkung von gestapelten Variablen auf die Tests:

4. Öffnen Sie das Tabellenerstellungsprogramm erneut (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
5. Ziehen Sie die Variable *Höchster Schulabschluss* aus der Variablenliste und legen Sie sie im Zeilenbereich unter *Geschlecht* ab.
6. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
			Keeping house	C E	C E		C	
Highest degree	LT High school		A C E					
	High school							
	Junior college	B		B		B		
	Bachelor	B				B		
	Graduate	B						

Abbildung 64. Vergleich der Spaltenanteile

Wenn *Höchster Schulabschluss* mit *Geschlecht* gestapelt ist, werden 19 Reihen von Spaltenmitteltests durchgeführt: die 14 bereits beschriebenen sowie eine für jede Ebene von *Höchster Schulabschluss*. Den Kategorien von *Familienstand* werden dieselben Buchstabenschlüssel zugewiesen.

Beachten Sie die folgenden Punkte bei den Tabellenergebnissen:

- Die Testergebnisse für die 14 bereits durchgeführten Testreihen sind identisch.
- Personen, die nur einen Hauptschulabschluss besitzen, sind unter den verwitweten Befragten häufiger anzutreffen als unter den verheirateten, geschiedenen oder ledigen.
- Personen mit einem Hochschulabschluss finden sich tendenziell häufiger unter den verheirateten, geschiedenen und ledigen Personen als unter verwitweten Personen.

Anmerkung zu Gewichtungen und Mehrfachantwortsets

Fallgewichtungen basieren immer auf Häufigkeiten, nicht auf Antworten, auch wenn es sich bei einer der Variablen um eine Mehrfachantwortvariable handelt.

Kapitel 11. Mehrfachantwortsets

In benutzerdefinierten Tabellen und der Diagrammerstellung kann eine besondere Art von "Variable" verwendet werden, die als **Mehrfachantwortset** bezeichnet wird. Bei Mehrfachantwortsets handelt es sich nicht um "Variablen" im üblichen Sinn. Mehrfachantwortsets können nicht im Dateneditor angezeigt werden, sie werden von anderen Prozeduren nicht erkannt. Mehrfachantwortsets verwenden mehrere Variablen, um Antworten auf Fragen aufzuzeichnen, auf welche der Befragte mehr als eine Antwort geben kann. Sie werden wie kategoriale Variablen behandelt und bieten weitestgehend dieselben Möglichkeiten wie kategoriale Variablen.

Mehrfachantwortsets werden aus mehreren Variablen in der Datendatei gebildet. Beim Mehrfachantwortset handelt es sich um ein spezielles Konstrukt innerhalb einer Datendatei. Mehrfachantwortsets können in einer Datendatei im IBM® SPSS Statistics-Format definiert und gespeichert werden. Das Importieren oder Exportieren von Mehrfachantwortsets in bzw. aus anderen Dateiformaten ist jedoch nicht möglich. Mit der Option "Dateneigenschaften kopieren" können Sie Mehrfachantwortsets aus anderen IBM SPSS Statistics-Datendateien kopieren. Sie finden diese Option im Menü "Daten" im Fenster "Dateneditor".

Beispieldatendatei

Für die Beispiele dieses Kapitels wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Häufigkeiten, Antworten, Prozentsätze und Gesamtwerte

Für Mehrfachantwortsets sind dieselben Auswertungsstatistiken verfügbar wie für kategoriale Variablen. Darüber hinaus stehen für Mehrfachantwortsets einige zusätzliche Statistiken zur Verfügung.

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

2. Ziehen Sie *Nachrichtenquellen*, die aussagekräftige Beschriftung für das Mehrfachantwortset \$*mltnews*, aus der Variablenliste in den Erstellungsbereich und legen Sie es im Bereich für die Zeilen ab.

Das Symbol neben der Variablen in der Variablenliste kennzeichnet die Variable als Set von dichotomen Variablen.



Abbildung 65. Symbol "Set von dichotomen Variablen"

Bei einem Set von dichotomen Variablen ist jede "Kategorie" tatsächlich eine separate Variable und die Kategoriebeschriftungen sind die Variablenbeschriftungen (bzw. die Variablennamen für Variablen, für die keine Variablenbeschriftungen definiert wurden). In diesem Beispiel repräsentieren die angezeigten Häufigkeiten die Anzahl der Fälle, in denen die Antwort für eine Variable im Set *Ja* lautete.

3. Klicken Sie in der Tabellenvorschau im Erstellungsbereich mit der rechten Maustaste auf *Nachrichtenquellen* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
4. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken, und klicken Sie anschließend auf **Zuweisen**.

5. Klicken Sie mit der rechten Maustaste erneut auf *Nachrichtenquellen* und wählen Sie im Popup-Menü **Auswertungsstatistik** aus.
6. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
7. Klicken Sie auf **Der Auswahl zuweisen** und anschließend auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %
News sources	Get news from internet	867	41.7%
	Get news from radio	551	26.5%
	Get news from television	1077	51.8%
	Get news from news magazines	294	14.1%
	Get news from newspapers	805	38.7%
	Total	2081	100.0%

Abbildung 66. Häufigkeiten und Spaltenprozentage in einem Set von dichotomen Variablen

Unstimmige Summen

Wenn Sie die Zahlen in der Tabelle betrachten, fällt Ihnen möglicherweise auf, dass ein relativ großer Unterschied zwischen dem Wert für "Gesamtsumme" und der angenommenen Summe aus den einzelnen Werten besteht. Dies gilt insbesondere, wenn die Gesamtsumme wesentlich niedriger ausfällt, als erwartet. Der Grund dafür besteht darin, dass die Anzahl für die einzelnen Kategorien in der Tabelle der Anzahl der Fälle mit dem Wert 1 (einer *Ja*-Antwort) für diese Variable entspricht und die Gesamtzahl der *Ja*-Antworten für alle fünf Variablen im Set von dichotomen Variablen leicht die Gesamtzahl der Fälle in der Datendatei übersteigen kann.

Demgegenüber wird die Gesamthäufigkeit aus der Summe der Fälle mit einer *Ja*-Antwort für mindestens eine Variable im Set gebildet und kann daher in keinem Fall die Gesamtzahl der Fälle in der Datendatei überschreiten. In diesem Beispiel ist die Gesamthäufigkeit von 2.081 fast 800 niedriger als die Gesamtzahl der Fälle in der Datendatei. Wenn keine Variable fehlende Werte enthält, bedeutet das, dass fast 800 Befragte angegeben haben, dass Sie aus keiner der angegebenen Quellen Nachrichten beziehen. Die Spaltenprozentage werden auf der Grundlage der Gesamthäufigkeit ermittelt. Daher werden die Spaltenprozentage in diesem Beispiel auf mehr als die 100 % summiert, die als Wert für die Gesamtspaltenprozentage angezeigt werden.

Stimmige Summen

Obwohl "Häufigkeiten" normalerweise ein relativ eindeutiger Begriff ist, zeigt das obige Beispiel, dass dennoch Verwirrung im Zusammenhang mit Gesamtsummen für Mehrfachantwortsets auftreten kann. Für Mehrfachantwortsets bezieht sich die tatsächlich gewünschte Auswertungsstatistik oftmals auf *Antworten*.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die Variable *Nachrichtenquellen* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
3. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Antworten** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
4. Wählen Sie in der Liste "Statistik" die Option **Antworten als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
5. Klicken Sie auf **Der Auswahl zuweisen** und anschließend auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	Responses	Column Responses %
News sources	Get news from internet	867	41.7%	867	24.1%
	Get news from radio	551	26.5%	551	15.3%
	Get news from television	1077	51.8%	1077	30.0%
	Get news from news magazines	294	14.1%	294	8.2%
	Get news from newspapers	805	38.7%	805	22.4%
	Total	2081	100.0%	3594	100.0%

Abbildung 67. Antworten und Antworten als Spaltenprozentsätze in einem Set von dichotomen Variablen

Für jede "Kategorie" in einem Set von dichotomen Variablen ist der Wert für *Antworten* identisch mit dem Wert für *Anzahl*. Dies ist immer der Fall bei Sets von dichotomen Variablen. Die Gesamtsummen können sich jedoch deutlich unterscheiden. Die Gesamtzahl der Antworten beträgt 3.594. Dieser Wert übersteigt die Gesamthäufigkeit um mehr als 1.500 und die Gesamtzahl der Fälle in der Datendatei um mehr als 700.

Bei den Prozentsätzen betragen die Gesamtwerte für *Anzahl als Spalten%* und *Antworten als Spalten%* beide 100 %. Für die einzelnen Kategorien im Set für dichotome Variablen sind die Prozentsätze bei den Antworten als Spaltenprozente jedoch wesentlich niedriger. Der Grund dafür liegt in der Tatsache, dass als Prozentbasis für die Antworten als Spaltenprozente die Gesamtzahl der Antworten verwendet wird, die im gegebenen Fall 3.594 beträgt. Dies führt zu wesentlich niedrigeren Prozentsätzen, als eine Berechnung auf der Grundlage des Werts 2.081, anhand dessen die Spaltenprozentsätze berechnet werden.

Die Gesamtsummen für die Prozentsätze sind größer als 100 %

Sowohl die Spaltenprozentsätze als auch die Werte für Antworten als Spaltenprozente erzielen in diesem Beispiel Gesamtprozentsätze von 100 %, obgleich sich die einzelnen Werte in der Spalte *Anzahl als Spalten%* auf deutlich mehr als 100 % summieren lassen. Wie können Sie also Prozentsätze auf der Grundlage der Gesamthäufigkeit und nicht anhand der Gesamtanzahl der Antworten darstellen und gleichzeitig dafür sorgen, dass der Prozentsatz für die "Gesamtsumme" präzise die Summe der Prozentsätze für die einzelnen Kategorien widerspiegelt?

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die Variable *Nachrichtenquellen* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
3. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Antworten als Spalten% (Basis: Anzahl)** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
4. Klicken Sie auf **Der Auswahl zuweisen** und anschließend auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news from newspapers	805	38.7%	805	22.4%	38.7%
	Total	2081	100.0%	3594	100.0%	172.7%

Abbildung 68. Berechnung der Werte für Antworten als Spaltenprozente auf der Grundlage der Prozentsätze für die Anzahl

Verwenden von Mehrfachantwortsets mit anderen Variablen

Generell werden Mehrfachantwortsets genauso verwendet wie kategoriale Variablen. Sie können beispielsweise ein Mehrfachantwortset zusammen mit einer kategorialen Variablen als Kreuztabelle darstellen lassen oder es innerhalb einer kategorialen Variablen verschachteln.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste in das Vorschauenfenster und legen Sie sie auf der linken Seite im Bereich für die Zeilen ab. Dadurch wird das Mehrfachantwortset *Nachrichtenquellen* innerhalb der Geschlechtskategorien verschachtelt.
3. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf *Geschlecht* und heben Sie im Popup-Menü die Auswahl von **Variablenbeschriftung anzeigen** auf.
4. Führen Sie denselben Schritt für *Nachrichtenquellen* aus.

Daraufhin werden die Spalten mit den Variablenbeschriftungen aus der Tabelle entfernt, da sie in diesem Fall nicht erforderlich sind.

5. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	375	41.9%	375	24.4%	41.9%
	Total	896	100.0%	1539	100.0%	171.8%
Female	Get news from internet	508	42.9%	508	24.7%	42.9%
	Get news from radio	318	26.8%	318	15.5%	26.8%
	Get news from television	626	52.8%	626	30.5%	52.8%
	Get news from news magazines	173	14.6%	173	8.4%	14.6%
	Get news from newspapers	430	36.3%	430	20.9%	36.3%
	Total	1185	100.0%	2055	100.0%	173.4%

Abbildung 69. Mehrfachantwortsets in einer kategorialen Variablen verschachtelt

Statistikquellenvariablen und verfügbare Auswertungsstatistiken

Wenn eine Tabelle keine metrische Variable enthält, werden kategoriale Variablen und Mehrfachantwortsets hinsichtlich der Statistikquellenvariablen identisch behandelt: Die Statistikquellenvariable ist die innerste verschachtelte Variable in der Quellendimension der Statistik. Da einige Auswertungsstatistiken ausschließlich nur Mehrfachantwortsets zugewiesen werden können, muss das Mehrfachantwortset die innerste verschachtelte Variable in der Statistikquellendimension bilden, wenn Sie eine spezielle Auswertungsstatistik für Mehrfachantworten erhalten möchten.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ändern Sie die Verschachtelungsreihenfolge, indem Sie die Variable *Nachrichtenquellen* aus der Tabellenvorschau im Erstellungsbereich ziehen und sie links von der Variablen *Geschlecht* ablegen.

Alle speziellen Auswertungsstatistiken für Mehrfachantworten, d. h. Antworten und Antworten als Spaltenprozente, werden aus der Tabellenvorschau entfernt, weil die kategoriale Variable *Geschlecht* nun die innerste verschachtelte Variable und damit die Quellenvariable der Statistik darstellt.

Glücklicherweise werden diese Einstellungen vorübergehend im Tabellenerstellungsprogramm gespeichert. Wenn Sie die Variable *Nachrichtenquellen* zur ursprünglichen Verschachtelung in der Variablen *Geschlecht* zurückverschieben, werden alle Auswertungsstatistiken im Zusammenhang mit Antworten in der Tabellenvorschau wiederhergestellt.

Sets von kategorialen Variablen und doppelte Antworten

Sets von kategorialen Variablen verfügen über eine Funktion, die bei Sets von dichotomen Variablen nicht zur Verfügung steht: die Fähigkeit Doppeltantworten zu zählen. In vielen Fällen handelt es sich bei doppelten Antworten in Sets von kategorialen Variablen vermutlich um Codierungsfehler. Wenn die Frage in einer Umfrage beispielsweise "Welche drei Nationen produzieren Ihrer Meinung nach die besten Autos?" lautet, ist die Antwort *Schweden, Deutschland und Schweden* wahrscheinlich ungültig.

In anderen Fällen können Doppeltantworten jedoch durchaus gültig sein. Auf die Frage "In welchen Ländern wurden Ihre letzten drei Autos produziert?" wäre die Antwort *Schweden, Deutschland und Schweden* beispielsweise völlig plausibel.

Benutzerdefinierte Tabellen bieten die Möglichkeit, doppelte Antworten in Sets von kategorialen Variablen zu berücksichtigen. In der Standardeinstellung werden doppelte Antworten nicht gezählt, Sie können jedoch angeben, dass Doppeltantworten in die Zählung aufgenommen werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen zu löschen.
3. Ziehen Sie die Variable *Autohersteller, aktuelle Kfz-Modelle* aus der Variablenliste in den Erstellungsbereich und legen Sie sie im Bereich für die Zeilen ab.

Das Symbol neben der Variablen in der Variablenliste kennzeichnet die Variable als Set von kategorialen Variablen.



Abbildung 70. Symbol "Set von kategorialen Variablen"

Bei Sets von kategorialen Variablen repräsentieren die angezeigten Kategorien die Wertbeschriftungen, die für alle Variablen im Set definiert sind. (Bei Sets von dichotomen Variablen stellen die Kategorien hingegen die Variablenbeschriftungen für die einzelnen Variablen im Set dar.)

4. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die Variable *Autohersteller, aktuelle Kfz-Modelle* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
5. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus, indem Sie darauf klicken, und klicken Sie anschließend auf **Zuweisen**.
6. Klicken Sie mit der rechten Maustaste erneut auf die Variable *Autohersteller, aktuelle Kfz-Modelle* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
7. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Antworten** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
8. Klicken Sie auf **Der Auswahl zuweisen** und anschließend auf **OK**, um die Tabelle zu erstellen.

		Count	Responses
Car maker, most recent cars	American	1938	1938
	Japanese	1327	1327
	Korean	695	695
	German	693	693
	Swedish	360	360
	Other	343	343
	Total	2832	5356

Abbildung 71. Set von kategorialen Variablen: Häufigkeiten und Antworten ohne Duplikate

In der Standardeinstellung werden doppelte Antworten nicht gezählt, daher sind die Werte in dieser Tabelle für die einzelnen Kategorien in den Spalten *Anzahl* und *Antworten* identisch. Lediglich die Gesamtsummen unterscheiden sich.

9. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
10. Klicken Sie auf die Registerkarte **Optionen**.
11. Wählen Sie die Option **Doppeltantworten für Sets von kategorialen Variablen zählen** aus, indem Sie darauf klicken.
12. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count	Responses
Car maker, most recent cars	American	1938	2797
	Japanese	1327	1717
	Korean	695	760
	German	693	754
	Swedish	360	383
	Other	343	359
	Total	2832	6770

Abbildung 72. Set von kategorialen Variablen mit Doppeltantworten

In dieser Tabelle besteht ein deutlicher Unterschied zwischen den Werten in der Spalte *Anzahl* und den Werten in der Spalte *Antworten*, insbesondere zwischen Werten für amerikanische Autos. Dies ist ein Hinweis darauf, dass viele Befragte Halter mehrerer amerikanischer Autos waren.

Signifikanztests bei Mehrfachantwortsets

Mehrfachantwortsets werden bei Signifikanztests im Wesentlichen wie kategoriale Variablen verwendet.

- Bei Tests auf Unabhängigkeit (Chi-Quadrat) oder Vergleichen der Spaltenanteile (Z-Tests) werden die Tests mit den Häufigkeiten durchgeführt. Die Häufigkeiten müssen daher eine der Auswertungsstatistiken sein, die in der Tabelle angezeigt werden.
- Bei Sets von kategorialen Variablen werden keine Test durchgeführt, in denen Spaltenanteile oder Spaltenmittel verglichen werden (T-Tests), wenn auf der Registerkarte "Optionen" die Option **Doppeltantworten für multiple Kategoriesets zählen** ausgewählt ist. Weitere Informationen finden Sie in „Benutzerdefinierte Tabellen: Registerkarte "Optionen"“ auf Seite 17.

Test auf Unabhängigkeit bei Mehrfachantwortsets

In diesem Beispiel wird eine Kreuztabelle einer kategorialen Variablen und eines Mehrfachantwortsets erstellt und für die Kreuztabelle wird ein Chi-Quadrat-Test auf Unabhängigkeit durchgeführt.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen zu löschen.
3. Ziehen Sie *Nachrichtenquellen*, die aussagekräftige Beschriftung für das Set von dichotomen Variablen *\$mltnews*, aus der Variablenliste in den Erstellungsbereich und legen Sie es im Bereich für die Spalten ab.
4. Ziehen Sie die Variable *Geschlecht* aus der Variablenliste und legen Sie sie im Spaltenbereich im Erstellungsbereich ab.
5. Klicken Sie auf die Registerkarte **Teststatistiken**.
6. Wählen Sie **Tests auf Unabhängigkeit (Chi-Quadrat)** aus.
7. Falls die Option **Mehrfachantwortvariablen in Tests einschließen** nicht bereits ausgewählt ist, wählen Sie sie aus.
8. Klicken Sie auf **OK**, um die Prozedur auszuführen.

		News sources
Gender	Chi-square	10.266
	df	5
	Sig.	.068

Abbildung 73. Ergebnisse des Chi-Quadrat-Tests

Das Signifikanzniveau von 0,068 für den Chi-Quadrat-Test deutet an, dass sich Männer und Frauen wahrscheinlich nicht signifikant darin unterscheiden, aus welchen Quellen sie ihre Nachrichten beziehen. (Hierbei wird angenommen, dass Sie einen Signifikanzwert von höchstens 0,05 als Kriterium zum Bestimmen der statistischen Signifikanz heranziehen.)

Vergleichen von Spaltenmitteln mit Mehrfachantwortsets

In diesem Beispiel werden die Mittelwerte einer metrischen Variablen innerhalb von Kategorien berechnet, die durch ein Mehrfachantwortset festgelegt werden. Außerdem wird jedes Kategoriemittel mit jedem anderen Kategoriemittel verglichen, um signifikante Unterschiede zu ermitteln.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen zu löschen.
3. Ziehen Sie *Nachrichtenquellen*, die aussagekräftige Beschriftung für das Set von dichotomen Variablen *\$mltnews*, aus der Variablenliste in den Erstellungsbereich und legen Sie es im Bereich für die Spalten ab.
4. Ziehen Sie *Alter der befragten Person* in den Erstellungsbereich und legen Sie es im Bereich für die Zeilen ab.
5. Klicken Sie auf die Registerkarte **Teststatistiken**.
6. Wählen Sie **Spaltenmittel vergleichen (T-Tests)** aus.
7. Falls die Option **Mehrfachantwortvariablen in Tests einschließen** nicht bereits ausgewählt ist, wählen Sie sie aus.
8. Klicken Sie auf **OK**, um die Prozedur auszuführen.

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	Mean	Mean	Mean	Mean	Mean
Age of respondent	52	40	48	40	40

Comparisons of Column Means

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	(A)	(B)	(C)	(D)	(E)
Age of respondent	B C D E		B D E		

Results are based on two-sided tests assuming equal variances with significance level 0.05. For each significant pair, the key of the smaller category appears under the category with larger mean.

Abbildung 74. Ergebnisse des Signifikanztests

- Jede Kategorie des Mehrfachantwortsets wird durch einen Buchstaben (A, B, C, D, E) bezeichnet. Für jede Kategorie, bei der der Mittelwert einer anderen Kategorie sowohl niedriger als der Mittelwert der vorliegenden Kategorie ist als auch signifikant von diesem abweicht, wird der Buchstabe der Kategorie mit dem niedrigeren Mittelwert angezeigt.

- *Nachrichten aus Zeitungen* (A) weist das höchste mittlere Alter auf und die Mittelwerte aller anderen Kategorien weichen signifikant von diesem Mittelwert ab.
- *Nachrichten aus dem Fernsehen* (C) weist das nächsthöchste mittlere Alter auf und die Mittelwerte aller weiteren Kategorien (B, D und E) weichen signifikant von diesem Mittelwert ab. (Wie bereits gezeigt, weicht C außerdem signifikant von A ab.)
- Die mittleren Alter von *Nachrichten aus Zeitschriften* (B), *Nachrichten aus dem Radio* (D) und *Nachrichten aus dem Internet* (E) weichen nicht signifikant voneinander ab.

Kapitel 12. Fehlende Werte

Viele Datendateien enthalten eine gewisse Menge an fehlenden Daten. Fehlende Daten können auf die verschiedensten Faktoren zurückgeführt werden. Möglicherweise beantworteten Teilnehmer an einer Umfrage nicht jede Frage, bestimmte Variablen sind für einige Fälle nicht geeignet und Codierfehler können dazu führen, dass einige Werte verworfen werden.

In IBM SPSS Statistics gibt es zwei Arten von fehlenden Werten:

- **Benutzerdefiniert fehlende Werte.** Werte, für die definiert wurde, dass sie fehlende Daten enthalten. Diesen Werten können Wertbeschriftungen zugewiesen werden, um anzuzeigen, warum die Daten fehlen (z. B. der Code 99 und die Wertbeschriftung *Nicht zutreffend* für Schwangerschaft bei Männern).
- **Systemdefiniert fehlende Werte.** Wenn für eine numerische Variable kein Wert vorhanden ist, wird ihr der systemdefiniert fehlende Wert zugewiesen. Dies wird durch einen Punkt in der Datenansicht des Dateneditors angegeben.

Es stehen mehrere Funktionen zur Verfügung, mit denen die Auswirkungen von fehlenden Daten kompensiert und sogar Muster in fehlenden Daten analysiert werden können. Dieses Kapitel hat jedoch ein weitaus einfacheres Ziel: Es wird erläutert, wie fehlende Daten in benutzerdefinierten Tabellen verarbeitet werden und welche Auswirkungen fehlende Daten auf die Berechnung von Auswertungsstatistiken haben.

Beispieldatendatei

Für die Beispiele in diesem Kapitel wird die Datendatei *missing_values.sav* verwendet. Weitere Informationen finden Sie in Datendateien. Hierbei handelt es sich um eine einfache, frei erfundene Datendatei mit lediglich einer Variablen und zehn Fällen, die nur dazu dienen soll, grundlegende Konzepte zu fehlenden Werten zu veranschaulichen.

Tabellen ohne fehlende Werte

In der Standardeinstellung werden benutzerdefiniert fehlende Kategorien in benutzerdefinierten Tabellen nicht angezeigt (systemdefiniert fehlende Werte werden generell nicht angezeigt).

1. Wählen Sie in den Menüs Folgendes aus:
Analysieren > Tabellen > Benutzerdefinierte Tabellen
2. Ziehen Sie im Tabellenerstellungsprogramm die *Variable mit fehlenden Werten* (die einzige Variable in der Datei) aus der Variablenliste in den Erstellungsbereich und legen Sie sie im Bereich für die Zeilen ab.
3. Klicken Sie mit der rechten Maustaste im Erstellungsbereich auf die Variable und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
4. Klicken Sie im Dialogfeld "Kategorien und Gesamtsummen" auf die Option **Gesamt** (aktivieren Sie diese) und klicken Sie anschließend auf **Zuweisen**.
5. Klicken Sie erneut mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die *Variable mit fehlenden Werten* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
6. Wählen Sie im Dialogfeld "Auswertungsstatistik" in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
7. Klicken Sie auf **Der Auswahl zuweisen**.

Möglicherweise werden Sie eine leichte Diskrepanz zwischen den in der Tabellenvorschau im Erstellungsbereich angezeigten Kategorien und den in der Liste "Kategorien" (unterhalb der Variablenliste auf der linken Seite des Dialogfelds des Tabellenerstellungsprogramms) feststellen. Die Liste "Kategorien" enthält eine Kategorie mit der Beschriftung *Fehlende Werte*, die in der Tabellenvorschau nicht ent-

halten ist, da Kategorien mit fehlenden Werten in der Standardeinstellung ausgeschlossen sind. Die Beschriftung enthält das Wort "Werte" im Plural. Dadurch wird angegeben, dass die Variable mindestens zwei benutzerdefiniert fehlende Kategorien aufweist.

8. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	Total	7	100.0%

Abbildung 75. Tabelle ohne fehlende Werte

Diese Tabelle ist anscheinend perfekt. Die Summen der Kategoriewerte ergeben die Gesamtsummen und die Prozentsätze spiegeln genau die Werte wider, die Sie erhalten würden, wenn Sie die Gesamtanzahl als Prozentbasis verwenden (z. B. $3/7 = 0,429$ oder 42,9 %). Die Gesamtanzahl entspricht jedoch nicht der Gesamtanzahl von Fällen in der Datendatei, sondern der Gesamtanzahl von Fällen mit **nicht fehlenden** Werten, also Fällen, die für diese Variable keine benutzerdefiniert fehlenden bzw. systemdefiniert fehlenden Werte aufweisen.

Aufnahmen von fehlenden Werten in Tabellen

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf *Variable mit fehlenden Werten* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
3. Klicken Sie im Dialogfeld "Kategorien und Gesamtsummen" auf die Option **Fehlende Werte** (aktivieren Sie diese) und klicken Sie anschließend auf **Zuweisen**.

Die Tabelle enthält jetzt eine Kategorie für *Fehlende Werte*. Obwohl in der Tabellenvorschau lediglich eine Kategorie für fehlende Werte angezeigt wird, werden in der Tabelle sämtliche benutzerdefiniert fehlenden Kategorien angezeigt.

4. Klicken Sie erneut mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die *Variable mit fehlenden Werten* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
5. Klicken Sie im Dialogfeld "Auswertungsstatistik" auf die Option **Angepasste Auswertungsstatistik für Gesamtsummen und Zwischenergebnisse** (aktivieren Sie diese).
6. Wählen Sie in der Liste "Statistik" für die angepasste Auswertungsstatistik die Option **Gültige N** und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
7. Wiederholen Sie diesen Vorgang für **Gesamtanzahl**.
8. Klicken Sie auf **Der Auswahl zuweisen** und anschließend im Tabellenerstellungsprogramm auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	Valid N	Total N
Variable with missing values	Low	2	22.2%		
	Medium	3	33.3%		
	High	2	22.2%		
	Don't know	1	11.1%		
	Not applicable	1	11.1%		
	Total	9	100.0%		

Abbildung 76. Tabelle mit fehlenden Werten

Die beiden benutzerdefiniert fehlenden Kategorien *Weiß nicht* und *Nicht zutreffend* werden nun in der Tabelle angezeigt, wobei die Gesamtsumme jetzt 9 statt 7 beträgt. Dies spiegelt die beiden zusätzlichen Fälle mit benutzerdefiniert fehlenden Werten wider (einer pro benutzerdefiniert fehlender Kategorie). Die Prozentsätze in den Spalten wurden ebenfalls geändert, da sie auf der Anzahl der nicht

fehlenden und benutzerdefiniert fehlenden Werte beruhen. Nur systemdefiniert fehlende Werte wurden bei der Berechnung der Prozentsätze nicht berücksichtigt.

Gültige N zeigt die Gesamtanzahl der nicht fehlenden Fälle (7) an und *Gesamtanzahl* zeigt die Gesamtanzahl der Fälle einschließlich der benutzerdefiniert fehlenden und systemdefiniert fehlenden Fälle an. Die Gesamtanzahl von Fällen beträgt 10, also ein Fall mehr als die Anzahl von nicht fehlenden und benutzerdefiniert fehlenden Werten, die in der Spalte *Anzahl* als Gesamtanzahl angezeigt werden. Die Ursache dafür ist, dass ein Fall mit einem systemdefiniert fehlenden Wert vorhanden ist.

9. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
10. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die *Variable mit fehlenden Werten* und wählen Sie im Popup-Menü die Option **Auswertungsstatistik** aus.
11. Wählen Sie in der oberen Liste "Statistik" (nicht die Liste im Gruppenfeld "Angepasste Auswertungsstatistik für Gesamtsummen und Zwischenergebnisse") die Option **Gültige N als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
12. Wiederholen Sie diesen Vorgang für **Gesamtanzahl als Spalten%**.
13. Sie können diese Auswertungen auch in die entsprechende Liste im Gruppenfeld "Angepasste Auswertungsstatistik für Gesamtsummen und Zwischenergebnisse" aufnehmen.
14. Klicken Sie auf **Der Auswahl zuweisen** und anschließend auf **OK**, um die Tabelle zu erstellen.

		Count	Column N %	Column Valid N %	Column Total N %	Valid N	Total N
Variable with missing values	Low	2	22.2%	28.6%	20.0%		
	Medium	3	33.3%	42.9%	30.0%		
	High	2	22.2%	28.6%	20.0%		
	Don't know	1	11.1%	.0%	10.0%		
	Not applicable	1	11.1%	.0%	10.0%		
	Total	9	100.0%	100.0%	100.0%		

Abbildung 77. Tabelle mit fehlenden Werten, gültigen Prozentsätzen und Gesamtprozentsätzen

- *Anzahl als Spalten%* gibt den Prozentsatz in jeder Kategorie auf der Grundlage der Anzahl nicht fehlender und benutzerdefiniert fehlender Werte an (da benutzerdefiniert fehlende Werte ausdrücklich in die Tabelle aufgenommen wurden).
- *Gültige N als Spalten%* gibt den Prozentsatz in jeder Kategorie ausschließlich auf der Grundlage der gültigen, nicht fehlenden Fälle an. Diese Werte entsprechen den Spaltenprozentsätzen in der ursprünglichen Tabelle, in der keine benutzerdefiniert fehlenden Werte enthalten waren.
- *Gesamtanzahl als Spalten%* gibt den Prozentsatz in jeder Kategorie auf der Grundlage aller Fälle einschließlich der Fälle mit benutzerdefiniert fehlenden und systemdefiniert fehlenden Werten an. Wenn Sie die einzelnen Kategorieprozentsätze in dieser Kategorie addieren, werden Sie feststellen, dass deren Summe lediglich 90 % beträgt, da ein Fall der insgesamt 10 Fälle (10 %) den systemdefiniert fehlenden Wert aufweist. Obwohl dieser Fall in der Berechnungsgrundlage für die Prozentsätze berücksichtigt wurde, wird in der Tabelle keine Kategorie mit systemdefiniert fehlenden Werten bereitgestellt.

Kapitel 13. Formatieren und Anpassen von Tabellen

Formatieren und Anpassen von Tabellen

Bei benutzerdefinierten Tabellen können Sie verschiedene Eigenschaften für die Tabellenformatierung bereits bei der Tabellenerstellung festlegen. Dazu gehören:

- Anzeigeformat und Beschriftungen für Auswertungsstatistiken
- Minimale und maximale Breite für Datenspalten
- Text oder Wert für die Anzeige in leeren Zellen

Diese Einstellungen werden in der Schnittstelle des Tabellenerstellungsprogramms beibehalten, bis Sie sie ändern, die Einstellungen für das Tabellenerstellungsprogramm zurücksetzen oder eine andere Datendatei öffnen. Auf diese Weise können Sie mehrere Tabellen mit übereinstimmender Formatierung erstellen, ohne sie nach dem Erstellen manuell bearbeiten zu müssen. Sie können diese Formatierungseinstellungen auch zusammen mit allen anderen Tabellenparametern speichern, indem Sie die Befehlssyntax mit der Schaltfläche "Einfügen" in der Schnittstelle des Tabellenerstellungsprogramms in ein Syntaxfenster einfügen. Anschließend können Sie die Syntax in einer Datei speichern.

Viele Formatierungseigenschaften von Tabellen können auch nach dem Erstellen geändert werden. Hierzu können Sie alle im Viewer für Pivot-Tabellen verfügbaren Formatierungsfunktionen verwenden. In diesem Kapitel wird jedoch hauptsächlich das Festlegen der Tabellenformatierung vor dem Erstellen der Tabelle behandelt. Weitere Informationen zu Pivot-Tabellen finden Sie im Hilfesystem. Geben Sie als Stichwort auf der Registerkarte "Index" Pivot-Tabellen ein.

Beispieldatendatei

Für die Beispiele dieses Kapitels wird die Datendatei *survey_sample.sav* verwendet. Weitere Informationen finden Sie in Datendateien.

In allen angegebenen Beispielen werden die Variablenbeschriftungen in Dialogfeldern in alphabetischer Reihenfolge angezeigt. Die Anzeigeeigenschaften für Variablenlisten werden im Dialogfeld "Optionen" auf der Registerkarte "Allgemein" festgelegt (Menü "Bearbeiten", Befehl "Optionen").

Anzeigeformat für Auswertungsstatistiken

In den benutzerdefinierten Tabellen wird versucht, den Auswertungsstatistiken möglichst geeignete Standardformate zuzuweisen. In manchen Fällen ist es jedoch sinnvoll, diese Formate durch andere zu ersetzen.

1. Wählen Sie in den Menüs Folgendes aus:

Analysieren > Tabellen > Benutzerdefinierte Tabellen

2. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.
3. Ziehen Sie die Variable *Vertrauen in das Fernsehen* unter *Alterskategorie* im Zeilenbereich und legen Sie sie dort ab. Dadurch werden die beiden Variablen in der Zeilendimension gestapelt.
4. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Alle Zeilenvariablen auswählen** aus.
5. Klicken Sie mit der rechten Maustaste erneut auf *Alterskategorie* und wählen Sie im Popup-Menü die Option **Kategorien und Gesamtsummen** aus.
6. Wählen Sie im Dialogfeld "Kategorien und Gesamtsummen" die Option **Gesamt** aus (aktivieren Sie diese) und klicken Sie anschließend auf **Zuweisen**.

7. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf eine der Variablen und wählen Sie im Voranzeige Popup-Menü die Option **Auswertungsstatistik** aus.
8. Wählen Sie in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
9. Wählen Sie die Option **Angepasste Auswertungsstatistik für Gesamtsummen und Zwischenergebnisse** aus (aktivieren Sie sie).
10. Wählen Sie in der Liste "Statistik" für die angepasste Auswertungsstatistik die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
11. Wiederholen Sie diesen Vorgang für **Mittelwert**.
12. Klicken Sie dann auf **Allen zuweisen**.

Die Platzhalterwerte in der Tabellenvorschau geben das Standardformat für jede Auswertungsstatistik wieder.

- Für Häufigkeiten lautet das Standardanzeigeformat **nnnn**, also Ganzzahlwerte ohne Dezimalstellen.
- Für Prozentsätze lautet das Standardanzeigeformat **nnnn.n%**, also Zahlen mit einer einzelnen Dezimalstelle und einem Prozentzeichen nach dem Wert.
- Für den Mittelwert ist das Standardanzeigeformat für die beiden Variablen *unterschiedlich*.

Bei Auswertungsstatistiken, die keine Form von Häufigkeit (z. B. Gültige N und Gesamtanzahl) oder Prozentsatz darstellen, wird als Standardanzeigeformat das Anzeigeformat verwendet, das für die Variable im Dateneditor definiert wurde. Wenn Sie die Variablen in der Variablenansicht des Dateneditors betrachten, werden Sie feststellen, dass *Alterskategorie* (Variable *alt_kat*) mit zwei Dezimalstellen definiert ist, während *Vertrauen in das Fernsehen* (Variable *ver_tv*) ohne (d. h. mit null) Dezimalstellen definiert ist.

Dies ist wahrscheinlich einer der Fälle, in denen das Standardformat nicht erwünscht ist, da beide Werte möglichst mit derselben Anzahl von Dezimalstellen angezeigt werden sollen.

13. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf eine der Variablen und wählen Sie im Voranzeige Popup-Menü die Option **Auswertungsstatistik** aus.
Für den Mittelwert wird in der Zelle "Format" der Liste "Anzeigen" das Format *Auto* gezeigt. Dies bedeutet, dass das für die Variable definierte Anzeigeformat verwendet wird und die Zelle "Dezimalstellen" inaktiviert ist. Um die Anzahl der Dezimalstellen festzulegen, müssen Sie zunächst ein anderes Format auswählen.
14. Klicken Sie in der Liste "Anzeigen" für die angepasste Auswertungsstatistik auf die Zelle "Format" für den Mittelwert und wählen Sie in der Dropdown-Liste der Formate die Option **nnnn** aus.
15. Geben Sie in der Zelle "Dezimalstellen" den Wert 1 ein.
16. Klicken Sie dann auf **Allen zuweisen**, um diese Einstellung beiden Variablen zu zuweisen.

Aus der Tabellenvorschau ist nun ersichtlich, dass beide Mittelwerte mit einer Dezimalstelle angezeigt werden. (Sie könnten diese Tabelle nun unmittelbar erstellen. Der "Mittelwert" für *Alterskategorie* ist jedoch möglicherweise etwas schwierig zu interpretieren, da die eigentlichen numerischen Codes für diese Variable nur im Bereich von 1 bis 6 liegen.)

Anzeigen von Beschriftungen für Auswertungsstatistiken

Neben den Anzeigeformaten für Auswertungsstatistiken können Sie auch die Beschriftungen für jede Auswertungsstatistik festlegen.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf **Zurücksetzen**, um frühere Einstellungen im Tabellenerstellungsprogramm zu löschen.
3. Ziehen Sie die Variable *Alterskategorie* aus der Variablenliste des Tabellenerstellungsprogramms in den Bereich "Zeilen" im Erstellungsbereich.

4. Ziehen Sie die Variable *Zahlungsart letzte Woche* aus der Variablenliste in den Erstellungsbereich und legen Sie sie im Spaltenbereich ab.
5. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die Variable *Alterskategorie* und wählen Sie im Kontextmenü die Option **Auswertungsstatistik** aus.
6. Wählen Sie in der Liste "Statistik" die Option **Anzahl als Spalten%** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
7. Doppelklicken Sie in der Liste "Anzeigen" in der Zelle "Beschriftung" auf das Wort *Spalte*, um den Inhalt der Zelle zu bearbeiten. Löschen Sie das Wort *Spalte* aus der Beschriftung und ersetzen Sie es durch %.
8. Bearbeiten Sie die Zelle "Beschriftung" für *Anzahl* auf entsprechende Weise und ersetzen Sie den Eintrag durch N.
Ändern Sie bei dieser Gelegenheit das Format der Statistik "Anzahl als Spalten%", um das nun überflüssige Prozentzeichen zu entfernen. (Die Spaltenbeschriftung gibt jetzt an, dass die Spalte Prozentsätze enthält.)
9. Klicken Sie auf die Zelle "Format" für *Anzahl als Spalten%* und wählen Sie in der Dropdown-Liste der Formate die Option **nnnn.n**.
10. Klicken Sie dann auf **Der Auswahl zuweisen**.
In der Tabellenvorschau werden das geänderte Anzeigeformat und die geänderten Beschriftungen angezeigt.
11. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Abbildung 78. Tabelle mit geänderten Beschriftungen für Auswertungsstatistiken

Spaltenbreite

Sie werden bemerkt haben, dass die Tabelle im vorherigen Beispiel recht breit ist. Eine Lösung für dieses Problem besteht darin, einfach die Zeilen- und Spaltenvariablen zu tauschen. Alternativ dazu können die Spaltenbreiten reduziert werden, da sie wesentlich breiter als notwendig erscheinen. (Die Beschriftungen für die Auswertungsstatistik wurden nämlich gekürzt, damit die Spalten schmaler gemacht werden können.)

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf die Registerkarte **Optionen**.
3. Wählen Sie im Gruppenfeld "Breite der Datenspalten" die Option **Anpassen**.
4. Geben Sie im Feld "Maximum" den Wert 36 ein. (Stellen Sie sicher, dass im Feld "Einheiten" die Option **Punkt** ausgewählt ist.)
5. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Abbildung 79. Tabelle mit reduzierten Spaltenbreiten

Die Tabelle ist nun wesentlich übersichtlicher.

Anzeigewert für leere Zellen

In der Standardeinstellung wird in leeren Zellen (Zellen, die keine Fälle enthalten) eine 0 angezeigt. In diesen Zellen kann jedoch auch nichts (d. h., die Zellen werden leer gelassen) oder eine Textzeichenfolge angezeigt werden.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Klicken Sie auf die Registerkarte **Optionen**.
3. Wählen Sie im Gruppenfeld "Darstellung der Datenzelle" neben "Leere Zellen" die Option **Text** und geben Sie **Kein** ein.
4. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	None	None	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	None	None	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Abbildung 80. Tabelle, bei der in leeren Zellen "Kein" angezeigt wird

In den vier leeren Zellen der Tabelle wird nun statt des Werts 0 der Text *Kein* angezeigt.

Anzeigewert für fehlende Statistiken

Wenn eine Statistik nicht berechnet werden kann, wird in der Standardeinstellung ein Punkt (.) angezeigt. Dies ist das Symbol, mit dem der systemdefiniert fehlende Werte angegeben wird. Dieser Wert unterscheidet sich von einer "leeren" Zelle und daher wird der Anzeigewert für fehlende Statistiken getrennt vom Anzeigewert für Zellen ohne Fälle festgelegt.

1. Öffnen Sie das Tabellenerstellungsprogramm (Menü "Analysieren", Untermenü "Tabellen", Befehl "Benutzerdefinierte Tabellen").
2. Ziehen Sie die Variable *Täglicher Fernsehkonsum in Stunden* aus der Variablenliste in den Erstellungsbereich und legen Sie am Anfang des Spaltenbereichs über *Zahlungsart letzte Woche* ab.
Da *Täglicher Fernsehkonsum in Stunden* eine metrische Variable ist, wird sie automatisch als Quellvariable der Statistik verwendet und die Auswertungsstatistik wird auf den Mittelwert gesetzt.
3. Klicken Sie mit der rechten Maustaste in der Tabellenvorschau im Erstellungsbereich auf die Variable *Täglicher Fernsehkonsum in Stunden* und wählen Sie im Kontextmenü die Option **Auswertungsstatistik** aus.
4. Wählen Sie in der Liste "Statistik" die Option **Gültige N** aus und klicken Sie auf die Pfeilschaltfläche, um sie in die Liste "Anzeigen" zu übertragen.
5. Klicken Sie auf **Der Auswahl zuweisen**.
6. Klicken Sie auf die Registerkarte **Optionen**.

7. Geben Sie im Textfeld "Statistiken, die nicht berechnet werden können" den Text NV ein.
8. Klicken Sie auf **OK**, um die Tabelle zu erstellen.

		Hours per day watching TV											
		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N
Age category	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8
	25 to 34	3	134	5	2	2	30	2	29	2	52	2	22
	35 to 44	3	136	2	5	3	30	2	34	2	47	3	46
	45 to 54	2	90	2	4	2	22	2	36	2	45	2	34
	55 to 64	3	40	NA	None	3	7	2	15	2	23	3	15
	65 or older	3	18	2	2	1	1	NA	0	1	2	3	11

Abbildung 81. Tabelle, bei der für fehlende Statistiken "NV" angezeigt wird

Der Text NV wird in drei Zellen der Tabelle als Mittelwert angezeigt. Der entsprechende Wert für Gültige N liefert jeweils die Erklärung hierfür: Es sind keine Fälle vorhanden, anhand derer der Mittelwert berechnet werden kann.

Sie werden jedoch eine leichte Diskrepanz bemerken: Einer der drei Werte für "Gültige N" wird nicht, wie für Zellen ohne Fälle anzunehmen, mit der Beschriftung *Kein*, sondern als 0 angezeigt. Die Ursache hierfür ist, dass zwar keine gültigen Fälle zum Berechnen des Mittelwerts vorhanden sind, die Kategorie jedoch nicht völlig leer ist. Wenn Sie nochmals die ursprüngliche Tabelle mit nur zwei kategorialen Variablen betrachten, werden Sie erkennen, dass in dieser Kategorie eigentlich drei Fälle vorhanden sind. Es sind jedoch keine gültigen Fälle vorhanden, da alle drei Fälle fehlende Werte für die metrische Variable *Täglicher Fernsehkonsum in Stunden* aufweisen.

Kapitel 14. Beispieldateien

Die zusammen mit dem Produkt installierten Beispieldateien finden Sie im Unterverzeichnis *Samples* des Installationsverzeichnisses. Für jede der folgenden Sprachen gibt es einen eigenen Ordner innerhalb des Unterverzeichnisses "Samples": Deutsch, Englisch, Französisch, Italienisch, Japanisch, Koreanisch, Polnisch, Russisch, Spanisch, Traditionelles Chinesisch und Vereinfachtes Chinesisch.

Nicht alle Beispieldateien stehen in allen Sprachen zur Verfügung. Wenn eine Beispieldatei nicht in einer Sprache zur Verfügung steht, enthält der jeweilige Sprachordner eine englische Version der Beispieldatei.

Beschreibungen

Im Folgenden finden Sie Kurzbeschreibungen der in den verschiedenen Beispielen in der Dokumentation verwendeten Beispieldateien.

- **accidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die alters- und geschlechtsabhängige Risikofaktoren für Autounfälle in einer bestimmten Region untersucht. Jeder Fall entspricht einer Kreuzklassifikation von Alterskategorie und Geschlecht.
- **adl.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die Vorteile einer vorgeschlagenen Therapieform für Schlaganfallpatienten zu ermitteln. Ärzte teilten weibliche Schlaganfallpatienten nach dem Zufallsprinzip jeweils einer von zwei Gruppen zu. Die erste Gruppe erhielt die physische Standardtherapie, die zweite erhielt eine zusätzliche Emotionstherapie. Drei Monate nach den Behandlungen wurden die Fähigkeiten der einzelnen Patienten, übliche Alltagsaktivitäten auszuführen, als ordinale Variablen bewertet.
- **advert.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Einzelhändlers geht, die Beziehungen zwischen den in Werbung investierten Beträgen und den daraus resultierenden Umsätzen zu untersuchen. Zu diesem Zweck hat er die Umsätze vergangener Jahre mit den zugehörigen Werbeausgaben zusammengestellt.
- **aflatoxin.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests von Maisernten auf Aflatoxin geht, ein Gift, dessen Konzentration stark zwischen und innerhalb von Ernteerträgen schwankt. Ein Getreideverarbeitungsbetrieb hat aus 8 Ernteerträgen je 16 Proben erhalten und das Aflatoxinniveau in Teilen pro Milliarde (PPB - Parts per Billion) gemessen.
- **anorectic.sav.** Bei der Ausarbeitung einer standardisierten Symptomatologie anorektischen/bulimischen Verhaltens führten Forscher¹ eine Studie mit 55 Jugendlichen mit bekannten Essstörungen durch. Jeder Patient wurde vier Mal über einen Zeitraum von vier Jahren untersucht, es fanden also insgesamt 220 Beobachtungen statt. Bei jeder Beobachtung erhielten die Patienten Scores für jedes von 16 Symptomen. Die Symptomscores fehlen für Patient 71 zum Zeitpunkt 2, Patient 76 zum Zeitpunkt 2 und Patient 47 zum Zeitpunkt 3, wodurch 217 gültige Beobachtungen verbleiben.
- **bankloan.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Bank geht, den Anteil der nicht zurückgezahlten Kredite zu reduzieren. Die Datei enthält Informationen zum Finanzstatus und demografischen Hintergrund von 850 früheren und potenziellen Kunden. Bei den ersten 700 Fällen handelt es sich um Kunden, denen bereits ein Kredit gewährt wurde. Bei den letzten 150 Fällen handelt es sich um potenzielle Kunden, deren Kreditrisiko die Bank als gering oder hoch einstufen möchte.
- **bankloan_binning.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen zum Finanzstatus und demografischen Hintergrund von 5.000 früheren Kunden enthält.

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

- **behavior.sav.** In einem klassischen Beispiel² wurden 52 Schüler/Studenten gebeten, die Kombinationen aus 15 Situationen und 15 Verhaltensweisen auf einer 10-Punkte-Skala von 0 = "ausgesprochen angemessen" bis 9 = "ausgesprochen unangemessen" zu bewerten. Die Werte werden über die einzelnen Personen gemittelt und als Unähnlichkeiten verwendet.
- **behavior_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine zweidimensionale Lösung für *behavior.sav*.
- **brakes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik geht, die Scheibenbremsen für Hochleistungsfahrzeuge herstellt. Die Datendatei enthält Messungen des Durchmessers von 16 Scheiben aus 8 Produktionsmaschinen. Der Zieldurchmesser für die Scheiben ist 322 Millimeter.
- **breakfast.sav.** In einer klassischen Studie³ wurden 21 MBA-Studenten der Wharton School mit ihren Lebensgefährten darum gebeten, 15 Frühstücksartikel in der Vorzugsreihenfolge von 1 = "am meisten bevorzugt" bis 15 = "am wenigsten bevorzugt" zu ordnen. Die Präferenzen wurden in sechs unterschiedlichen Szenarios erfasst, von "Overall preference" (Allgemein bevorzugt) bis "Snack, with beverage only" (Imbiss, nur mit Getränk).
- **breakfast-overall.sav.** Diese Datei enthält die Daten zu den bevorzugten Frühstücksartikeln, allerdings nur für das erste Szenario, "Overall preference" (Allgemein bevorzugt).
- **broadband_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die die Anzahl der Abonnenten eines Breitbandservice, nach Region geordnet, enthält. Die Datendatei enthält die monatlichen Abbonnentenzahlen für 85 Regionen über einen Zeitraum von vier Jahren.
- **broadband_2.sav** Diese Datendatei stimmt mit *broadband_1.sav* überein, enthält jedoch Daten für weitere drei Monate.
- **car_insurance_claims.sav.** Ein an anderer Stelle⁴ vorgestelltes und analysiertes Dataset bezieht sich auf Schadensansprüche für Autos. Die durchschnittliche Höhe der Schadensansprüche lässt sich mit Gamma-Verteilung modellieren. Dazu wird eine inverse Verknüpfungsfunktion verwendet, um den Mittelwert der abhängigen Variablen mit einer linearen Kombination aus Alter des Versicherungsnehmers, Fahrzeugtyp und Fahrzeualter in Bezug zu setzen. Die Anzahl der eingereichten Schadensansprüche kann als Skalierungsgewichtung verwendet werden.
- **car_sales.sav.** Diese Datendatei enthält hypothetische Verkaufsschätzungen, Listenpreise und physische Spezifikationen für verschiedene Fahrzeugfabrikate und -modelle. Die Listenpreise und physischen Spezifikationen wurden von *edmunds.com* und Herstellerwebsites entnommen.
- **car_sales_uprepared.sav.** Hierbei handelt es sich um eine modifizierte Version der Datei *car_sales.sav*, die keinerlei transformierte Versionen der Felder enthält.
- **carpet.sav.** In einem bekannten Beispiel⁵ möchte ein Unternehmen einen neuen Teppichreiniger vermarkten und dazu den Einfluss von fünf Faktoren auf die Präferenz durch den Verbraucher untersuchen: Verpackungsgestaltung, Markenname, Preis, Gütesiegel *Good Housekeeping* und Geld-zurück-Garantie. Die Verpackungsgestaltung setzt sich aus drei Faktorebenen zusammen, die sich durch die Position der Auftragebürste unterscheiden. Außerdem gibt es drei Markennamen (*K2R*, *Glory* und *Bis-sell*), drei Preisstufen sowie je zwei Ebenen (Nein oder Ja) für die letzten beiden Faktoren. 10 Kunden stufen 22 Profile ein, die durch diese Faktoren definiert sind. Die Variable *Preference* enthält den Rang der durchschnittlichen Einstufung für die verschiedenen Profile. Ein niedriger Rang bedeutet eine starke Präferenz. Diese Variable gibt ein Gesamtmaß der Präferenz für die Profile an.
- **carpet_prefs.sav.** Diese Datendatei beruht auf denselben Beispielen, wie für *carpet.sav* beschrieben, enthält jedoch die tatsächlichen Einstufungen durch jeden der 10 Kunden. Die Kunden wurden gebeten, die 22 Produktprofile in der Reihenfolge ihrer Präferenzen einzustufen. Die Variablen *PREF1* bis *PREF22* enthalten die IDs der zugeordneten Profile, wie in *carpet_plan.sav* definiert.

2. Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

3. Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

4. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

5. Green, P. E., and Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **catalog.sav.** Diese Datendatei enthält hypothetische monatliche Verkaufszahlen für drei Produkte, die von einem Versandhaus verkauft werden. Daten für fünf mögliche Prädiktorvariablen wurden ebenfalls aufgenommen.
- **catalog_seasfac.sav.** Diese Datendatei ist mit *catalog.sav* identisch, außer, dass ein Set von saisonalen Faktoren, die mithilfe der Prozedur "Saisonale Zerlegung" berechnet wurden, sowie die zugehörigen Datumsvariablen hinzugefügt wurden.
- **cellular.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Mobiltelefonunternehmens geht, die Kundenabwanderung zu verringern. Propensity-Scores für die Abwanderungsneigung (von 0 bis 100) werden auf die Kunden angewendet. Kunden mit einem Score von 50 oder höher streben vermutlich einen Anbieterwechsel an.
- **ceramics.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Herstellers geht, der ermitteln möchte, ob ein neue, hochwertige Keramiklegierung eine größere Hitzebeständigkeit aufweist als eine Standardlegierung. Jeder Fall entspricht einem Test einer der Legierungen; die Temperatur, bei der das Keramikwälzlager versagte, wurde erfasst.
- **cereal.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Umfrage geht, bei der 880 Personen nach ihren Frühstücksgewohnheiten befragt wurden. Außerdem wurden Alter, Geschlecht, Familienstand und Vorliegen bzw. Nichtvorliegen eines aktiven Lebensstils (auf der Grundlage von mindestens zwei Trainingseinheiten pro Woche) erfasst. Jeder Fall entspricht einem Befragten.
- **clothing_defects.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Bekleidungsfabrik geht. Aus jeder in der Fabrik produzierten Charge entnehmen die Kontrolleure eine Stichprobe an Bekleidungsartikeln und zählen die Anzahl der Bekleidungsartikel die inakzeptabel sind.
- **coffee.sav.** Diese Datendatei enthält Daten zum wahrgenommenen Image von sechs Eiskaffee⁶. Bei den 23 Attributen des Eiskaffee-Image sollten die Teilnehmer jeweils alle Marken auswählen, die durch dieses Attribut beschrieben werden. Die sechs Marken werden als "AA", "BB", "CC", "DD", "EE" und "FF" bezeichnet, um Vertraulichkeit zu gewährleisten.
- **contacts.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Kontaktlisten einer Gruppe von Vertretern geht, die Computer an Unternehmen verkaufen. Die einzelnen Kontaktpersonen werden anhand der Abteilung, in der sie in ihrem Unternehmen arbeiten und anhand ihrer Stellung in der Unternehmenshierarchie in Kategorien eingeteilt. Außerdem werden der Betrag des letzten Verkaufs, die Zeit seit dem letzten Verkauf und die Größe des Unternehmens, in dem die Kontaktperson arbeitet, aufgezeichnet.
- **creditpromo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Kaufhauses geht, die Wirksamkeit einer kürzlich durchgeführten Kreditkartenwerbeaktion einzuschätzen. Dazu wurden 500 Karteninhaber nach dem Zufallsprinzip ausgewählt. Die Hälfte erhielt eine Werbebeilage, die einen reduzierten Zinssatz für Einkäufe in den nächsten drei Monaten ankündigte. Die andere Hälfte erhielt eine Standardwerbebeilage.
- **customer_dbase.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, das die Informationen in seinem Data Warehouse nutzen möchte, um spezielle Angebote für Kunden zu erstellen, die mit der größten Wahrscheinlichkeit darauf ansprechen. Nach dem Zufallsprinzip wurde ein Subset des Kundenstamms ausgewählt. Diese Kunden erhielten die speziellen Angebote und die Reaktionen wurden aufgezeichnet.
- **customer_information.sav.** Eine hypothetische Datendatei mit Kundenmailing-Daten wie Name und Adresse.
- **customer_subset.sav.** Ein Subset von 80 Fällen aus der Datei *customer_dbase.sav*.
- **debate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die paarige Antworten auf eine Umfrage unter den Zuhörern einer politischen Debatte enthält (Antworten vor und nach der Debatte). Jeder Fall entspricht einem Befragten.

6. Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

- **debate_aggregate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der die Antworten aus *debate.sav* aggregiert wurden. Jeder Fall entspricht einer Kreuzklassifikation der favorisierten Politiker vor und nach der Debatte.
- **demo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Kundendatenbank geht, die zum Zwecke der Zusendung monatlicher Angebote erworben wurde. Neben verschiedenen demografischen Informationen ist erfasst, ob der Kunde auf das Angebot geantwortet hat.
- **demo_cs_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den ersten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einer anderen Stadt. Außerdem sind IDs für Region, Provinz, Landkreis und Stadt erfasst.
- **demo_cs_2.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den zweiten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einem anderen Stadtteil aus den im ersten Schritt ausgewählten Städten. Außerdem sind IDs für Region, Provinz, Landkreis, Stadt, Stadtteil und Wohneinheit erfasst. Die Informationen zur Stichprobenziehung aus den ersten beiden Stufen des Stichprobenplans sind ebenfalls enthalten.
- **demo_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfrageinformationen enthält, die mit einem komplexen Stichprobendesign erfasst wurden. Jeder Fall entspricht einer anderen Wohneinheit. Es sind verschiedene Informationen zum demografischen Hintergrund und zur Stichprobenziehung erfasst.
- **diabetes_costs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen enthält, die von einer Versicherungsgesellschaft für an Diabetes erkrankte Versicherungsnehmer verwaltet werden. Jeder Fall entspricht einem anderen Versicherungsnehmer.
- **dietstudy.sav.** Diese hypothetische Datendatei enthält die Ergebnisse einer Studie der "Stillman-Diät"⁷. Jeder Fall entspricht einem Teilnehmer und enthält dessen Gewicht vor und nach der Diät in amerikanischen Pfund sowie mehrere Messungen des Triglyceridspiegels (in mg/100 ml).
- **dmdata.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen über Demografie und Einkäufe für ein Direktmarketing-Unternehmen enthält. *dmdata2.sav* enthält Informationen für ein Subset von Kontakten, die eine Testsendung erhalten, und *dmdata3.sav* enthält Informationen zu den verbleibenden Kontakten, die keine Testsendung erhalten.
- **dvdplayer.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Entwicklung eines neuen DVD-Players geht. Mithilfe eines Prototyps hat das Marketingteam Zielgruppendaten erfasst. Jeder Fall entspricht einem befragten Benutzer und enthält demografische Daten zu dem Benutzer sowie dessen Antworten auf Fragen zum Prototyp.
- **german_credit.sav.** Diese Daten sind aus dem Dataset "German credit" im Repository of Machine Learning Databases⁸ an der Universität von Kalifornien in Irvine entnommen.
- **grocery_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *grocery_coupons.sav*, wobei die wöchentlichen Einkäufe zusammengefasst sind, sodass jeder Fall einem anderen Kunden entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und der verzeichnete ausgegebene Betrag ist nun die Summe der Beträge, die in den vier Wochen der Studie ausgegeben wurden.
- **grocery_coupons.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfragedaten enthält, die von einer Lebensmittelkette erfasst wurden, die sich für die Kaufgewohnheiten ihrer Kunden interessiert. Jeder Kunde wird über vier Wochen beobachtet und jeder Fall entspricht einer Kundenwoche und enthält Informationen zu den Geschäften, in denen der Kunde einkauft sowie zu anderen Merkmalen, beispielsweise welcher Betrag in der betreffenden Woche für Lebensmittel ausgegeben wurde.
- **guttman.sav.** Bell⁹ legte eine Tabelle zur Darstellung möglicher sozialer Gruppen vor. Guttman¹⁰ verwendete einen Teil dieser Tabelle, bei der fünf Variablen, die Aspekte beschreiben, wie soziale Interakti-

7. Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

on, das Gefühl der Gruppenzugehörigkeit, die physische Nähe der Mitglieder und die Formalität der Beziehung, mit sieben theoretischen sozialen Gruppen gekreuzt wurden: "crowds" (Menschenmassen, beispielsweise die Zuschauer eines Fußballspiels), "audience" (Zuhörerschaften, beispielsweise die Personen im Theater oder bei einer Vorlesung), "public" (Öffentlichkeit, beispielsweise Zeitungsleser oder Fernsehzuschauer), "mobs" (Mobs, wie Menschenmassen, jedoch mit wesentlich stärkerer Interaktion), "primary groups" (Primärgruppen, vertraulich), "secondary groups" (Sekundärgruppen, freiwillig) und "modern community" (die moderne Gesellschaft, ein lockerer Zusammenschluss, der aus einer engen physischen Nähe und dem Bedarf an spezialisierten Dienstleistungen entsteht).

- **health_funding.sav.** Hierbei handelt es sich um eine hypothetische Datei, die Daten zur Finanzierung des Gesundheitswesens (Betrag pro 100 Personen), Krankheitsraten (Rate pro 10.000 Personen der Bevölkerung) und Besuche bei medizinischen Einrichtungen/Ärzten (Rate pro 10.000 Personen der Bevölkerung) enthält. Jeder Fall entspricht einer anderen Stadt.
- **hivassay.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Bemühungen eines pharmazeutischen Labors, einen Schnelltest zur Erkennung von HIV-Infektionen zu entwickeln. Die Ergebnisse des Tests sind acht kräftiger werdende Rotschattierungen, wobei kräftigeren Schattierungen auf eine höhere Infektionswahrscheinlichkeit hindeuten. Bei 2.000 Blutproben, von denen die Hälfte mit HIV infiziert war, wurde ein Labortest durchgeführt.
- **hourlywagedata.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zum Stundenlohn von Pflegepersonal in Praxen und Krankenhäusern mit unterschiedlich langer Berufserfahrung.
- **insurance_claims.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die ein Modell zur Kennzeichnung verdächtiger, potenziell betrügerischer Ansprüche erstellen möchte. Jeder Fall entspricht einem Anspruch.
- **insure.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die die Risikofaktoren untersucht, die darauf hinweisen, ob ein Kunde die Leistungen einer mit einer Laufzeit von 10 Jahren abgeschlossenen Lebensversicherung in Anspruch nehmen wird. Jeder Fall in der Datendatei entspricht einem Paar von Verträgen, je einer mit Leistungsforderung und der andere ohne, wobei die beiden Versicherungsnehmer in Alter und Geschlecht übereinstimmen.
- **judges.sav.** Hierbei handelt es sich um eine hypothetische Datendatei mit den Wertungen von ausgebildeten Kampfrichtern (sowie eines Sportliebhabers) zu 300 Kunstturnleistungen. Jede Zeile stellt eine Leistung dar; die Kampfrichter bewerteten jeweils dieselben Leistungen.
- **kinship_dat.sav.** Rosenberg und Kim¹¹ haben 15 Bezeichnungen für den Verwandtschaftsgrad untersucht (Tante, Bruder, Cousin, Tochter, Vater, Enkelin, Großvater, Großmutter, Enkel, Mutter, Nefte, Nichte, Schwester, Sohn, Onkel). Die beiden Analytiker baten vier Gruppen von College-Studenten (zwei weibliche und zwei männliche Gruppen), diese Bezeichnungen auf der Grundlage der Ähnlichkeiten zu sortieren. Zwei Gruppen (eine weibliche und eine männliche Gruppe) wurden gebeten, die Bezeichnungen zweimal zu sortieren; die zweite Sortierung sollte dabei nach einem anderen Kriterium erfolgen als die erste. So wurden insgesamt sechs "Quellen" erzielt. Jede Quelle entspricht einer Ähnlichkeitsmatrix mit 15 × 15 Elementen. Die Anzahl der Zellen ist dabei gleich der Anzahl der Personen in einer Quelle minus der Anzahl der gemeinsamen Platzierungen der Objekte in dieser Quelle.
- **kinship_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine dreidimensionale Lösung für *kinship_dat.sav*.
- **kinship_var.sav.** Diese Datendatei enthält die unabhängigen Variablen *gender* (Geschlecht), *gener* (Generation) und *degree* (Verwandtschaftsgrad), die zur Interpretation der Dimensionen einer Lösung für *kinship_dat.sav* verwendet werden können. Insbesondere können sie verwendet werden, um den Lösungsraum auf eine lineare Kombination dieser Variablen zu beschränken.
- **marketvalues.sav.** Diese Datendatei betrifft Hausverkäufe in einem Neubaugebiet in Algonquin, Illinois, in den Jahren 1999–2000. Diese Verkäufe sind in Grundbucheinträgen dokumentiert.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

11. Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

- **nhis2000_subset.sav.** Die "National Health Interview Survey (NHIS)" ist eine große, bevölkerungsbezogene Umfrage unter der US-amerikanischen Zivilbevölkerung. Es werden persönliche Interviews in einer landesweit repräsentativen Stichprobe von Haushalten durchgeführt. Für die Mitglieder jedes Haushalts werden demografische Informationen und Beobachtungen zum Gesundheitsverhalten und Gesundheitsstatus eingeholt. Diese Datendatei enthält ein Subset der Informationen aus der Umfrage des Jahres 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Datendatei und Dokumentation öffentlich zugänglich. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Zugriff erfolgte 2003.
- **ozone.sav.** Die Daten enthalten 330 Beobachtungen zu sechs meteorologischen Variablen zur Vorhersage der Ozonkonzentration aus den übrigen Variablen. Bei früheren Untersuchungen^{12, 13} fanden Wissenschaftler einige Nichtlinearitäten unter diesen Variablen, die die Standardverfahren bei der Regression behindern.
- **pain_medication.sav.** Diese hypothetische Datendatei enthält die Ergebnisse eines klinischen Tests für ein entzündungshemmendes Medikament zur Schmerzbehandlung bei chronischer Arthritis. Von besonderem Interesse ist die Zeitdauer, bis die Wirkung des Medikaments einsetzt und wie es im Vergleich mit bestehenden Medikamenten abschneidet.
- **patient_los.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen zu Patienten, die wegen des Verdachts auf Herzinfarkt in das Krankenhaus eingeliefert wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **patlos_sample.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen für eine Stichprobe von Patienten, denen während der Behandlung eines Herzinfarkts Thrombolytika verabreicht wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **poll_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die öffentliche Unterstützung für einen Gesetzentwurf zu ermitteln, bevor er im Parlament eingebracht wird. Die Fälle entsprechen registrierten Wählern. Für jeden Fall sind County, Gemeinde und Wohnviertel des Wählers erfasst.
- **poll_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *poll_cs.sav* aufgeführten Wähler. Die Stichprobe wurde gemäß dem in der Plandatei *poll_csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Beachten Sie jedoch Folgendes: Da im Stichprobenplan die PPS-Methode (Probability Proportional To Size - Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*poll_jointprob.sav*). Die zusätzlichen Variablen zum demografischen Hintergrund der Wähler und ihrer Meinung zum vorgeschlagenen Gesetzentwurf wurden nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **property_assess.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen Bezirk (County) zuständigen Immobilienbewerbers geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien, die im vergangenen Jahr in dem betreffenden County verkauft wurden. Jeder Fall in der Datendatei enthält die Gemeinde, in der sich die Immobilie befindet, den Bewerter, der die Immobilie besichtigt hat, die seit dieser Bewertung verstrichene Zeit, den zu diesem Zeitpunkt ermittelten Wert sowie den Verkaufswert der Immobilie.
- **property_assess_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen US-Bundesstaat zuständigen Immobilienbewerbers geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien in dem betreffenden Bundesstaat. Jeder Fall in der Datendatei enthält das County, die Gemeinde und das Wohnviertel, in dem sich die Immobilie befindet, die seit der letzten Bewertung verstrichene Zeit sowie den zu diesem Zeitpunkt ermittelten Wert.

12. Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

- **property_assess_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *property_assess_cs.sav* aufgeführten Immobilien. Die Stichprobe wurde gemäß dem in der Plandatei *property_assess_csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Die zusätzliche Variable *Current value* (Aktueller Wert) wurde nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **recidivism.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem früheren Straftäter und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Zeit bis zu seiner zweiten Festnahme, sofern diese innerhalb von zwei Jahren nach der ersten Festnahme erfolgte.
- **recidivism_cs_sample.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem früheren Straftäter, der im Juni 2003 erstmals aus der Haft entlassen wurde, und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Daten zu seiner zweiten Festnahme, sofern diese bis Ende Juni 2006 erfolgte. Die Straftäter wurden aus per Stichprobenziehung ermittelten Polizeidirektionen ausgewählt (gemäß dem in *recidivism_cs.plan* angegebenen Stichprobenplan). Da hierbei eine PPS-Methode (Probability Proportional To Size - Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav.** Eine hypothetische Datendatei mit Kauftransaktionsdaten wie Kaufdatum, gekauften Artikeln und Geldbetrag für jede Transaktion.
- **salesperformance.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Evaluierung von zwei neuen Verkaufsschulungen geht. 60 Mitarbeiter, die in drei Gruppen unterteilt sind, erhalten jeweils eine Standardschulung. Zusätzlich erhält Gruppe 2 eine technische Schulung und Gruppe 3 eine Praxisschulung. Die einzelnen Mitarbeiter wurden am Ende der Schulung einem Test unterzogen und der erzielte Score wurde erfasst. Jeder Fall in der Datendatei stellt einen Lehrgangsteilnehmer dar und enthält die Gruppe, der der Lehrgangsteilnehmer zugeteilt wurde sowie der von ihm in der Prüfung erreichte Score.
- **satisf.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Zufriedenheitsumfrage, die von einem Einzelhandelsunternehmen in 4 Filialen durchgeführt wurde. Insgesamt wurden 582 Kunden befragt. Jeder Fall gibt die Antworten eines einzelnen Kunden wieder.
- **screws.sav.** Diese Datendatei enthält Informationen zu den Eigenschaften von Schrauben, Bolzen, Muttern und Reißnägeln¹⁴.
- **shampoo_ph.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik für Haarpflegeprodukte geht. In regelmäßigen Zeitintervallen werden Messwerte von sechs separaten Ausgangschargen erhoben und ihr pH-Wert erfasst. Der Zielbereich ist 4,5-5,5.
- **ships.sav.** Ein an anderer Stelle¹⁵ vorgestelltes und analysiertes Dataset bezieht sich auf die durch Wellen verursachten Schäden an Frachtschiffen. Die Vorfalshäufigkeiten können unter Angabe von Schiffstyp, Konstruktionszeitraum und Betriebszeitraum gemäß einer Poisson-Rate modelliert werden. Das Aggregat der Betriebsmonate für jede Zelle der durch die Kreuzklassifizierung der Faktoren gebildeten Tabelle gibt die Werte für die Risikoanfälligkeit an.
- **site.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, neue Standorte für die betriebliche Expansion auszuwählen. Das Unternehmen beauftragte zwei Berater unabhängig voneinander mit der Bewertung der Standorte. Neben einem umfassenden Bericht gaben die Berater auch eine zusammenfassende Wertung für jeden Standort als "good" (gut) "fair" (mittelmäßig) oder "poor" (schlecht) ab.
- **smokers.sav.** Diese Datendatei wurde aus der Umfrage "National Household Survey of Drug Abuse" aus dem Jahr 1998 abstrahiert und stellt eine Wahrscheinlichkeitsstichprobe US-amerikanischer Haus-

14. Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

15. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

halte dar. (<http://dx.doi.org/10.3886/ICPSR02934>) Daher sollte der erste Schritt bei der Analyse dieser Datendatei darin bestehen, die Daten entsprechend den Bevölkerungstrends zu gewichten.

- **stocks.sav.** Diese hypothetische Datendatei umfasst Börsenkurse und -volumina für ein Jahr.
- **stroke_clean.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozeduren in der Option "Data Preparation" bereinigt wurde.
- **stroke_invalid.sav.** Diese hypothetische Datendatei enthält den ursprünglichen Zustand einer medizinischen Datenbank, der mehrere Dateneingabefehler aufweist.
- **stroke_survival.sav.** In dieser hypothetischen Datendatei geht es um die Überlebenszeiten von Patienten, die nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls mit einer Reihe von Problemen zu kämpfen haben. Nach dem Schlaganfall werden das Auftreten von Herzinfarkt, ischämischen Schlaganfall und hämorrhagischem Schlaganfall sowie der Zeitpunkt des Ereignisses aufgezeichnet. Die Stichprobe ist auf der linken Seite abgeschnitten, da sie nur Patienten enthält, die bis zum Ende des Rehabilitationsprogramms, das nach dem Schlaganfall durchgeführt wurde, überlebten.
- **stroke_valid.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozedur "Daten validieren" überprüft wurde. Sie enthält immer noch potenziell anomale Fälle.
- **survey_sample.sav.** Diese Datendatei enthält Umfragedaten einschließlich demografischer Daten und verschiedener Meinungskennzahlen. Sie beruht auf einem Subset der Variablen aus der NORC General Social Survey aus dem Jahr 1998. Allerdings wurden zu Demonstrationszwecken einige Daten abgeändert und weitere fiktive Variablen hinzugefügt.
- **tcm_kpi.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die die Werte der wöchentlichen wesentlichen Leistungsindikatoren für ein Unternehmen enthält. Sie enthält darüber hinaus wöchentliche Daten für zahlreiche steuerbare Metriken über denselben Zeitraum.
- **tcm_kpi_upd.sav.** Diese Datendatei stimmt mit *tcm_kpi.sav* überein, enthält jedoch Daten für weitere vier Wochen.
- **telco.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Telekommunikationsunternehmens geht, die Kundenabwanderung zu verringern. Jeder Fall entspricht einem Kunden und enthält verschiedene Informationen zum demografischen Hintergrund und zur Servicenutzung.
- **telco_extra.sav.** Diese Datendatei ähnelt der Datei *telco.sav*, allerdings wurden die Variablen "tenure" und die Log-transformierten Variablen zu den Kundenausgaben entfernt und durch standardisierte Log-transformierte Variablen ersetzt.
- **telco_missing.sav.** Diese Datendatei ist ein Subset der Datendatei *telco.sav*, allerdings wurde ein Teil der demografischen Datenwerte durch fehlende Werte ersetzt.
- **testmarket.sav.** Diese hypothetische Datendatei bezieht sich auf die Pläne einer Fast-Food-Kette, einen neuen Artikel in ihr Menü aufzunehmen. Es gibt drei mögliche Kampagnen zur Verkaufsförderung für das neue Produkt. Daher wird der neue Artikel in Filialen in mehreren zufällig ausgewählten Märkten eingeführt. An jedem Standort wird eine andere Form der Verkaufsförderung verwendet und die wöchentlichen Verkaufszahlen für das neue Produkt werden für die ersten vier Wochen aufgezeichnet. Jeder Fall entspricht einer Standortwoche.
- **testmarket_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *testmarket.sav*, wobei die wöchentlichen Verkaufszahlen zusammengefasst sind, sodass jeder Fall einem Standort entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren und die verzeichneten Verkaufszahlen sind nun die Summe der Verkaufszahlen während der vier Wochen der Studie.
- **tree_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_credit.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält.
- **tree_missing_data.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält und eine große Anzahl fehlender Werte aufweist.

- **tree_score_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_textdata.sav.** Eine einfache Datendatei mit nur zwei Variablen, die vor allem den Standardzustand von Variablen vor der Zuweisung von Messniveau und Wertbeschriftungen zeigen soll.
- **tv-survey.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Studie, die von einem Fernsehstudio durchgeführt wurde, das überlegt, ob die Laufzeit eines erfolgreichen Programms verlängert werden soll. 906 Personen wurden gefragt, ob sie das Programm unter verschiedenen Bedingungen ansehen würden. Jede Zeile entspricht einem Befragten; jede Spalte entspricht einer Bedingung.
- **ulcer_recurrence.sav.** Diese Datei enthält Teilinformationen aus einer Studie zum Vergleich der Wirksamkeit zweier Therapien zur Vermeidung des Wiederauftretens von Geschwüren. Es stellt ein gutes Beispiel für intervallzensierte Daten dar und wurde an anderer Stelle¹⁶ vorgestellt und analysiert.
- **ulcer_recurrence_recoded.sav.** In dieser Datei sind die Daten aus *ulcer_recurrence.sav* so umstrukturiert, dass das Modell der Ereigniswahrscheinlichkeit für jedes Intervall der Studie berechnet werden kann und nicht nur die Ereigniswahrscheinlichkeit am Ende der Studie. Sie wurde an anderer Stelle¹⁷ vorgestellt und analysiert.
- **verd1985.sav.** Diese Datendatei enthält eine Umfrage¹⁸. Die Antworten von 15 Subjekten auf 8 Variablen wurden aufgezeichnet. Die relevanten Variablen sind in drei Sets unterteilt. Set 1 umfasst *age* und *marital*, Set 2 besteht aus *pet* und *news* und in Set 3 finden sich *music* und *live*. Die Variable *pet* wird mehrfach nominal skaliert und die Variable *age* ordinal. Alle anderen Variablen werden einzeln nominal skaliert.
- **virus.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Internet-Service-Providers geht, der die Auswirkungen eines Virus auf seine Netze ermitteln möchte. Dabei wurde vom Moment der Virusentdeckung bis zu dem Zeitpunkt, zu dem die Virusinfektion unter Kontrolle war, der (ungefähre) prozentuale Anteil infizierter E-Mail in den Netzen erfasst.
- **wheeze_steubenville.sav.** Hierbei handelt es sich um ein Subset der Daten aus einer Langzeitstudie zu den gesundheitlichen Auswirkungen der Luftverschmutzung auf Kinder¹⁹. Die Daten enthalten wiederholte binäre Messungen des Keuchens von Kindern aus Steubenville, Ohio, im Alter von 7, 8, 9 und 10 Jahren sowie eine unveränderliche Angabe, ob die Mutter im ersten Jahr der Studie rauchte oder nicht.
- **workprog.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einem Arbeitsprogramm der Regierung, das versuchte, benachteiligten Personen bessere Arbeitsplätze zu verschaffen. Eine Stichprobe potenzieller Programmteilnehmer wurde beobachtet. Von diesen Personen wurden nach dem Zufallsprinzip einige für die Teilnahme an dem Programm ausgewählt. Jeder Fall entspricht einem Programmteilnehmer.
- **worldsales.sav.** Diese hypothetische Datendatei enthält Verkaufserlöse nach Kontinent und Produkt.

16. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

Bemerkungen

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

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

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

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

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

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

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

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

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
USA*

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

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

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

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

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

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

COPYRIGHTLIZENZ:

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

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

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

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

Marken

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

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

- Angepasste Auswertungsstatistik für Gesamtsummen 60
- Anzahl
 - und gültige N 64
- Anzeigeformate 26
 - Auswertungsstatistiken in benutzerdefinierten Tabellen 12, 99
- Anzeigen von Kategoriewerten 61
- Ausblenden der Beschriftungen für die Statistiken in benutzerdefinierten Tabellen 23
- Ausschließen von Kategorien
 - benutzerdefinierte Tabellen 12, 28
- Auswertungsstatistik 57
- Auswertungsstatistiken
 - Ändern des Beschriftungstexts 100
 - angepasste Auswertungsstatistik für Gesamtsummen 60
 - Anzeigeformat 99
 - gestapelte Tabellen 59
 - Quellendimension 57
 - Quellenvariable 57
 - unterschiedliche Auswertungen für unterschiedliche Variablen in gestapelten Tabellen 66

B

- Beispieldateien
 - Speicherort 105
- Benutzerdefiniert fehlende Werte 95
- Benutzerdefinierte Tabellen
 - Ändern der Beschriftung für die Auswertungsstatistik 24
 - Ändern der Dimension von Auswertungsstatistiken 11
 - Ändern des Messniveaus 1
 - angepasste Gesamtsummen 11
 - Anzeigeformate 6
 - Anzeigeformate für Auswertungsstatistiken 12
 - Ausblenden der Beschriftung für die Statistik 23
 - Ausblenden der Kategorien für das Zwischenergebnis 44
 - Ausschließen von fehlenden Werten bei der Auswertung metrischer Variablen 17
 - Ausschließen von Kategorien 12, 28
 - Auswertungsstatistik 7, 8, 9
 - berechnete Kategorien 12, 15, 47
 - Comperimeter-Tabellen 17, 53
 - Drucken von Tabellen mit Schichten 37
 - Eckenbeschriftungen 19
 - Einblenden und Ausblenden von Variablenamen und -beschriftungen 6
 - einfache Tabellen für kategoriale Variablen 23

- Benutzerdefinierte Tabellen (*Forts.*)
 - Erstellen einer Tabelle 4
 - Gesamtsummen 12, 25, 39
 - Gesamtsummen in Tabellen mit aus-
geschlossenen Kategorien 28
 - Häufigkeitstabellen 17, 53
 - kategoriale Variablen 1
 - Kompaktansicht 34
 - Kreuztabelle 25
 - leere Zellen 17
 - Mehrfachantwortsets 1, 87
 - metrische Variablen 1
 - Postcompute-Kategorien 15, 47
 - Prozentsätze 7, 8, 24, 26
 - Prozentsätze für Mehrfachantwort-
sets 9
 - Quellendimension der Statistik 26
 - Randsummen 27
 - Reduzieren von Kategorien 44
 - Schichtvariablen 35, 36, 37
 - Sets von kategorialen Variablen 17
 - Signifikanztests und Mehrfachantworten 85
 - Sortieren von Kategorien 28
 - Spaltenbreite 17
 - Stapeln von Variablen 31
 - Steuern der angezeigten Dezimalstellen 6
 - Tabellen für Mittelwert und Häufigkeit 11
 - Tabellen für Variablen mit gemeinsamen Kategorien 17, 53
 - Teststatistiken 19, 73
 - Titel 19
 - Titelzeilen 19
 - Umordnen von Kategorien 12
 - Verarbeitung aufgeteilter Dateien 4
 - Verschachteln von Schichtvariablen 37
 - verschachtelte Variablen 32, 34
 - Vertauschen von Zeilen- und Spaltenvariablen 35
 - Wertbeschriftungen für kategoriale Variablen 1
 - Zeilen- und Spaltenprozentsätze 24
 - Zwischenergebnisse 12, 39
- Berechnete Kategorien
 - Anzeigeformate 16
 - aus Zwischenergebnissen 48
 - Ausblenden von Kategorien im Ausdruck 48
 - benutzerdefinierte Tabellen 15, 47
- Bereich
 - benutzerdefinierte Tabellen 9
- Beschriftungen
 - Ändern des Beschriftungstexts für Auswertungsstatistiken 100

C

- Chi-Quadrat
 - benutzerdefinierte Tabellen 73
- Comperimeter-Tabellen 17, 53

D

- Datum
 - Einfügen des aktuellen Datums in benutzerdefinierte Tabellen 19
- Dezimalstellen
 - Steuern der in benutzerdefinierten Tabellen angezeigten Dezimalstellen 6, 26, 99

E

- Eckenbeschriftungen
 - benutzerdefinierte Tabellen 19

F

- Fehlende Werte 64, 95
 - Auswirkungen auf die Berechnung von Prozentsätzen 96
 - in benutzerdefinierte Tabellen aufnehmen 96

G

- Gesamtanzahl 96
- Gesamtsummen
 - Anzeigeposition 40
 - ausgeschlossene Kategorien 40
 - benutzerdefinierte Tabellen 12, 25, 39
 - Gesamtsummen für Gruppen 41
 - Randsummen in benutzerdefinierten Tabellen 27
 - Schichten 42
 - verschachtelte Tabellen 41
- Gesamtsummen für Gruppen 41
- Gesamtsummen für Untergruppen 41
- Gruppierte Auswertungen
 - metrische Variablen 67
- Gültige Anzahl
 - benutzerdefinierte Tabellen 9
- Gültige Werte 64, 96

H

- Häufigkeitstabellen
 - benutzerdefinierte Tabellen 17, 53

K

- Konfidenzintervalle 71
- Kreuztabelle
 - benutzerdefinierte Tabellen 25

L

- Leere Zellen
 - angezeigter Wert in benutzerdefinierten Tabellen 17, 102
- Löschen von Kategorien
 - benutzerdefinierte Tabellen 12, 28

M

- Maximum
 - benutzerdefinierte Tabellen 9
- Median 64
 - benutzerdefinierte Tabellen 9
- Mehrfachantwortsets 87
 - Doppeltantworten in Sets von kategorialen Variablen 17
 - Prozentsätze 9
 - Signifikanztests 85, 92, 93
- Messniveau
 - in benutzerdefinierten Tabellen ändern 1
- Metrische Variablen
 - Auswertungen gruppiert nach kategorialen Zeilen- und Spaltenvariablen 67
 - Auswertungsstatistik 63
 - gruppierte Auswertungen 67
 - mehrere Auswertungsstatistiken 64
 - Stapeln 63
 - Verschachtelung 68
- Minimum
 - benutzerdefinierte Tabellen 9
- Mittelwert 63
 - benutzerdefinierte Tabellen 9
- Modus
 - benutzerdefinierte Tabellen 9

P

- Postcompute-Kategorien
 - benutzerdefinierte Tabellen 15, 47
- Prozentsätze
 - fehlende Werte 96
 - in benutzerdefinierten Tabellen 7, 8, 24, 26
 - Mehrfachantwortsets 9

Q

- Quellenvariable für Auswertungsstatistik
 - metrische Variablen 68

R

- Reduzieren von Kategorien
 - benutzerdefinierte Tabellen 44

S

- Schichtvariablen
 - benutzerdefinierte Tabellen 35, 36, 37
 - Drucken von Tabellen mit Schichten 37
 - Stapeln von Schichtvariablen 36

Schichtvariablen (*Forts.*)

- Verschachteln von Schichtvariablen 37
- Signifikanztests
 - benutzerdefinierte Tabellen 19
 - Mehrfachantwortsets 92, 93
- Sortieren von Kategorien
 - benutzerdefinierte Tabellen 28
- Spaltenanteilestatistik
 - benutzerdefinierte Tabellen 79
- Spaltenbreite
 - Steuern in benutzerdefinierten Tabellen 17, 101
- Spaltenmittelstatistik
 - benutzerdefinierte Tabellen 76
- Standardabweichung
 - benutzerdefinierte Tabellen 9
- Stapeln von Variablen
 - benutzerdefinierte Tabellen 31
 - mehrere Quellenvariablen für Auswertungsstatistiken 59
 - metrische Variablen 63
 - Stapeln von Schichtvariablen 36
 - unterschiedliche Auswertungsstatistiken für unterschiedliche Variablen 66
- Statistik
 - angepasste Auswertungsstatistik für Gesamtsummen 60
 - Auswertungsstatistik 57
 - gestapelte Tabellen 59
 - Steuern der angezeigten Dezimalstellen 26
- Summe
 - benutzerdefinierte Tabellen 9
- Systemdefiniert fehlende Werte 95

T

- Tabellen
 - benutzerdefinierte Tabellen 1
- Tabellen für Mittelwert und Häufigkeit 11, 60
- Tabellen mit Schichten 37
- Teststatistiken
 - benutzerdefinierte Tabellen 19, 73
- Titel
 - benutzerdefinierte Tabellen 19
- Titelzeilen
 - benutzerdefinierte Tabellen 19

U

- Uhrzeit
 - Einfügen der aktuellen Uhrzeit in benutzerdefinierte Tabellen 19
- Umordnen von Kategorien
 - benutzerdefinierte Tabellen 12
- Unterschiedliche Auswertungsstatistiken für unterschiedliche Variablen
 - gestapelte Tabellen 66

V

- Variablen verschachteln
 - metrische Variablen 68

Variablenbeschriftungen

- Unterdrücken der Anzeige in benutzerdefinierten Tabellen 6
- Varianz
 - benutzerdefinierte Tabellen 9
- Verarbeitung aufgeteilter Dateien
 - benutzerdefinierte Tabellen 4
- Verschachtelte Variablen
 - benutzerdefinierte Tabellen 32, 34

W

- Weglassen von Kategorien
 - benutzerdefinierte Tabellen 28
- Werte
 - Anzeigen von Kategoriebeschriftungen und -werten 61
- Werte und Wertbeschriftungen 61

Z

- Zwischenergebnisse 43
 - Ausblenden der Kategorien für das Zwischenergebnis 44
 - benutzerdefinierte Tabellen 12, 39

