

# SPSS 22

Einführung in die moderne Datenanalyse

14., aktualisierte Auflage

Achim Bühl

**EXTRAS  
ONLINE**

ALWAYS LEARNING

PEARSON

# SPSS 22

Einführung in die moderne Datenanalyse

14., aktualisierte Auflage

Achim Bühl

Die Werte -1 und +1 können nur bei einer quadratischen Kreuztabelle angenommen werden.

### Kendalls Tau-c

Hier können die Werte -1 und +1 bei jeder Tabelle erreicht werden:

$$\tau_c = \frac{2 \cdot m \cdot (P - I)}{N^2 \cdot (m - 1)}$$

Dabei ist N die Gesamtsumme der Häufigkeiten; m ist die kleinere der Anzahlen der Zeilen und Spalten.

### 9.3.5 Weitere Assoziationsmaße

SPSS bietet weitere spezielle Assoziationsmaße an, die im Folgenden vorgestellt werden sollen.

#### Eta

Dieser Koeffizient ist geeignet, wenn die abhängige Variable Intervall- und die unabhängige Variable Ordinal- oder Nominalskalenniveau hat. Das Quadrat von Eta ist der Anteil der Gesamtvarianz, der durch die unabhängige Variable erklärt wird.

#### Kappa-Koeffizient

Cohens Kappa-Koeffizient kann nur für quadratische Kreuztabellen berechnet werden, in denen dieselben numerischen Kodierungen für die Zeilen- und die Spaltenvariable verwendet wurden. Der typische Anwendungsfall ist der, dass Personen oder Objekte durch zwei Gutachter beurteilt werden. Kappa gibt dann den Grad der Übereinstimmung zwischen den beiden Beurteilern an.

#### Risiko

Unter dieser Option verbergen sich bei SPSS drei verschiedene Koeffizienten, die für Vierfeldertafeln bestimmt werden können, wobei eine bestimmte Testsituation gegeben sein muss: Eine so genannte Risikovariable, die angibt, ob ein bestimmtes Ereignis eintrifft oder nicht, wird in Abhängigkeit von einer unabhängigen (ursächlichen) und ebenfalls dichotomen Variablen untersucht.

Dies soll an einem typischen Beispiel erläutert werden. Bei insgesamt 294 Probanden aus einer Studie über Depressionen ergab sich die folgende Häufigkeitsverteilung:

		Depression	
		Ja	Nein
Weiblich	Ja	a = 40	b = 143
	Männlich	c = 10	d = 101

Die Depression mit den beiden Kategorien ja – nein ist die Risikovariable, das Geschlecht mit den beiden Kategorien weiblich – männlich die unabhängige (ursächliche) Variable.

In Zusammenhang mit solchen Studien spricht man von Kohorten- bzw. Fall-Kontrollstudien. Dabei untersuchen Kohortenstudien eine festgelegte Fallgruppe, bei denen das untersuchte Ereignis noch nicht eingetreten ist, über einen gewissen Zeitraum hinweg und stellen fest, bei welchen Fällen dieses Ereignis eintritt und bei welchen nicht und ob sich das Risiko des Eintretens zwischen den Kategorien einer unabhängigen Variablen unterscheidet. Bei Fall-Kontrollstudien hingegen wird eine Fallgruppe, bei der das Ereignis bereits eingetreten ist, mit einer Kontrollgruppe verglichen.

Zwei der von SPSS berechneten Koeffizienten werden üblicherweise bei Kohortenstudien, der dritte Koeffizient bei Fall-Kontrollstudien verwendet. In Kohortenstudien wird für die beiden Kategorien der unabhängigen Variablen (hier: das Geschlecht) die Inzidenzrate bestimmt. Bei den weiblichen Probanden ist die Inzidenzrate für das Auftreten einer Depression

$$\frac{40}{40 + 143} = 0,219$$

Bei den männlichen Probanden ist die Inzidenzrate

$$\frac{10}{10 + 101} = 0,09$$

Der Quotient aus den beiden Inzidenzraten

$$\frac{0,219}{0,090} = 2,426$$

wird als relatives Risiko bezeichnet. Das Risiko, depressiv zu werden, liegt bei den Frauen um das 2,426-Fache über dem bei den Männern. Da der Rechner nicht wissen kann, welche der beiden Kodierungen der Risikovariablen für das Vorhandensein der Depression steht, wird dieses relative Risiko für beide Ausprägungen der Risikovariablen berechnet.

Bei Fall-Kontrollstudien wird eine etwas andere Variante der Koeffizientenberechnung, die auch Odds Ratio genannt wird, verwendet. So sind die »Chancen« (odds) bei den Frauen, depressiv zu werden, 40/143, bei den Männern 10/101. Das Chancenverhältnis (Odds Ratio) ist demnach

$$\frac{40 \cdot 101}{143 \cdot 10} = 2,825$$

Bezeichnen wir die vier Häufigkeiten der Vierfeldertafel mit a, b, c und d (siehe oben), so werden demnach von SPSS die drei folgenden Koeffizienten berechnet:

$$R0 = \frac{a \cdot d}{b \cdot c}$$

$$R1 = \frac{a \cdot (c + d)}{c \cdot (a + b)}$$

$$R2 = \frac{b \cdot (c + d)}{d \cdot (a + b)}$$

Wir wollen das gegebene Beispiel in SPSS durchrechnen.

- Laden Sie die Datei depr.sav.

Die Datei enthält als Risikovariable die Variable depr mit den beiden Kodierungen 1 = ja und 2 = nein sowie die unabhängige (ursächliche) Variable sex mit den beiden Kodierungen 1 = weiblich und 2 = männlich. Ferner ist als weitere Variable die Häufigkeitsvariable n enthalten.

- Treffen Sie die Menüwahl

*Daten*

*Fälle gewichten...*

und geben Sie n als Häufigkeitsvariable an.

- Definieren Sie anschließend in der Dialogbox *Kreuztabellen* die Variable sex als Zeilen- und die Variable depr als Spaltenvariable und aktivieren Sie über den Schalter *Statistiken...* die Option *Risiko*.

Es werden die folgenden Ergebnisse angezeigt.

	Wert	95%-Konfidenzintervall	
		Unterer	Oberer
Odds-Verhältnis für Geschlecht (Weiblich / Männlich)	2,825	1,350	5,911
Für Kohort Depression = Ja	2,426	1,265	4,655
Für Kohort Depression = Nein	,859	,780	,946
Anzahl der gültigen Fälle	294		

Es werden nacheinander Odds Ratio (R0) und die beiden relativen Risiken (R1 und R2) ausgegeben, wobei auch ein 95%-Konfidenzintervall berechnet wird.

Möchten Sie Odds Ratio und das relative Risiko korrekt berechnen, müssen Sie Folgendes beachten:

- ▶ Definieren Sie die ursächliche (unabhängige) Variable als Zeilenvariable und die Risikovariable als Spaltenvariable.
- ▶ In der ersten der beiden Zeilen der Vierfeldertafel muss die Gruppe mit dem größeren Risiko eingetragen sein.
- ▶ In der ersten der beiden Spalten der Vierfeldertafel muss die Kodierung für das Eintreffen des Ereignisses enthalten sein.

### Chi-Quadrat-Test nach McNemar

Der Chi-Quadrat-Test nach McNemar wird bei zwei abhängigen dichotomen Variablen eingesetzt; er ist in Kapitel 13.2.3 beschrieben.

### Cochran- und Mantel-Haenszel-Statistik

Diese Statistik knüpft an die Berechnung des Odds Ratio bei Vierfeldertafeln an, die mit der Option *Risiko* ermöglicht wird. Dabei kommt eine Schichtenvariable (Kovariate) ins Spiel, wobei getestet wird, ob sich das Quotenverhältnis (Odds Ratio) über die Kategorien dieser Variable hinweg signifikant von 1 (oder auch einem anderen Wert) unterscheidet. Ein Beispiel mag dies verdeutlichen.

- Laden Sie die Datei angst.sav.

Von insgesamt 1737 Personen sind drei Variablen gespeichert, nämlich das Geschlecht (1 = weiblich, 2 = männlich), eine Angabe darüber, ob eine Angststörung vorliegt (1 = ja, 2 = nein), und eine Angabe über das Körpergewicht (1 = nicht übergewichtig, 2 = übergewichtig). Wir wollen getrennt nach Übergewichtigen und Nichtübergewichtigen eine Kreuztabelle zwischen dem Geschlecht und dem Auftreten von Angststörungen erstellen und dabei jeweils das Odds Ratio berechnen.

- Treffen Sie die Menüwahl

#### *Daten*

##### *Aufgeteilte Datei...*

- Aktivieren Sie die Option *Ausgabe nach Gruppen aufteilen* und geben Sie die Variable gewicht als Gruppierungsvariable an.
- Wählen Sie anschließend aus dem Menü

#### *Analysieren*

##### *Deskriptive Statistiken*

##### *Kreuztabellen...*

- Tragen Sie die Variable sex als Zeilenvariable und die Variable angst als Spaltenvariable ein.
- Aktivieren Sie über den Schalter *Zellen...* die Ausgabe zeilenweiser Prozentwerte und über den Schalter *Statistiken...* die Option *Risiko*.

Die wesentliche Ausgabe ist im Folgenden wiedergegeben.

**Übergewicht = Nein**

**Übergewicht = Nein**

**Kreuztabelle Geschlecht\*Angststörung<sup>a</sup>**

		Angststörung		Gesamtsumme
		Ja	Nein	
Geschlecht	Weiblich	Anzahl	154	592
		% in Geschlecht	20,6%	79,4%
Männlich		Anzahl	79	715
		% in Geschlecht	9,9%	90,1%
Gesamtsumme		Anzahl	233	1307
		% in Geschlecht	15,1%	84,9%
				1540
				100,0%

a. Übergewicht = Nein

**Risikoschätzung<sup>a</sup>**

	Wert	95%-Konfidenzintervall	
		Unterer	Oberer
Odds-Verhältnis für Geschlecht (Weiblich / Männlich)	2,354	1,758	3,154
Für Kohort Angststörung = Ja	2,075	1,612	2,670
Für Kohort Angststörung = Nein	,881	,844	,920
Anzahl der gültigen Fälle	1540		

a. Übergewicht = Nein

**Übergewicht = Ja****Kreuztabelle Geschlecht \* Angststörung<sup>a</sup>**

		Angststörung		Gesamtsumme
		Ja	Nein	
Geschlecht	Weiblich	Anzahl	22	62
		% in Geschlecht	26,2%	73,8%
	Männlich	Anzahl	9	104
		% in Geschlecht	8,0%	92,0%
Gesamtsumme		Anzahl	31	166
		% in Geschlecht	15,7%	84,3%
				197
				100,0%

a. Übergewicht = Ja

**Risikoschätzung<sup>a</sup>**

	Wert	95%-Konfidenzintervall	
		Unterer	Oberer
Odds-Verhältnis für Geschlecht (Weiblich / Männlich)	4,100	1,776	9,468
Für Kohort Angststörung = Ja	3,288	1,597	6,771
Für Kohort Angststörung = Nein	,802	,698	,921
Anzahl der gültigen Fälle	197		

a. Übergewicht = Ja

Das Auftreten der Angststörung ist in beiden Fällen bei den Frauen deutlich erhöht. Das Odds Ratio beträgt bei den Nichtübergewichtigen 2,354 und bei den Übergewichtigen 4,100.

Wir wollen nun die Cochran- und Mantel-Haenszel-Statistik berechnen.

- Um zunächst die Aufteilung nach Gruppen wieder rückgängig zu machen, aktivieren Sie nach der Menüwahl

**Daten****Aufgeteilte Datei...**

die Option Alle Fälle analysieren, keine Gruppen bilden.

- In der Dialogbox Kreuztabellen geben Sie zusätzlich die Variable gewicht als Schichtvariable an, deaktivieren über den Schalter Statistiken... die Option Risiko und aktivieren die Option Cochran- und Mantel-Haenszel-Statistik.
- Unter der Option Gemeinsames Quoten-Verhältnis belassen Sie es bei der Voreinstellung von 1.

Wir beschränken uns darauf, die Ausgabe der Cochran- und Mantel-Haenszel-Statistik wiederzugeben.

#### Tests auf Homogenität des Odds-Verhältnisses

	Chi-Quadrat	df	Asymp. Sig. (zweiseitig)
Breslow-Day	1,522	1	,217
Tarone	1,522	1	,217

#### Tests auf bedingte Unabhängigkeit

	Chi-Quadrat	df	Asymp. Sig. (zweiseitig)
Cochran	44,665	1	,000
Mantel-Haenszel	43,724	1	,000

Unter Annahme der bedingten Unabhängigkeit ist die Cochran-Statistik als Chi-Quadrat-Verteilung mit 1 Freiheitsgrad nur asymptotisch verteilt, wenn die Anzahl der Schichten fest ist. Dagegen ist die Mantel-Haenszel-Statistik unter dieser Annahme immer als Chi-Quadrat-Verteilung mit 1 Freiheitsgrad asymptotisch verteilt. Beachten Sie, dass die Kontinuitätskorrektur aus der Mantel-Haenszel-Statistik entfernt wird, wenn die Summe der Differenzen zwischen beobachteten und erwarteten Größen null sind.

#### Schätzung des gemeinsamen Odds-Verhältnisses nach Mantel-Haenszel

Schätzung		2,503
In(Schätzung)		,918
Standardfehler von In(Schätzung)		,141
Asymp. Sig. (zweiseitig)		,000
Asymp. 95%-Konfidenzintervall	Gemeinsames Odds-Verhältnis	Untergrenze 1,901
		Obergrenze 3,297
	In(gemeinsames Odds-Verhältnis)	Untergrenze ,642
		Obergrenze 1,193

Schätzung des gemeinsamen Odds-Verhältnisses nach Mantel-Haenszel ist unter dem gemeinsamen Odds-Verhältnis der Voraussetzung 1,000 asymptotisch normal verteilt. Dies gilt auch für den natürlichen Logarithmus der Schätzung.

Die Ergebnisse nach Cochran und nach Mantel-Haenszel sind einander sehr ähnlich; in beiden Fällen ergibt sich über beide Gewichtsgruppen hinweg ein höchst signifikanter Unterschied ( $p < 0,001$ ) des Quotenverhältnisses (Odds Ratio) zum Wert 1. Sowohl nach Breslow-Day als auch nach Tarone kann die Annahme der Homogenität des Quotenverhältnisses über die beiden Gewichtsgruppen beibehalten werden ( $p = 0,217$ ).

Die Schätzung des gemeinsamen Quotenverhältnisses ergibt die gleichen Werte, die sich über die Option Risiko ergeben würden, wenn Sie keine Aufspaltung nach der Schichtenvariablen vorgenommen hätten.

# Analyse von Mehrfachantworten

In diesem Kapitel wollen wir die Kodierung und Analyse von Mehrfachantworten behandeln. Fragen mit Mehrfachantwortmöglichkeit spielen bei den meisten Fragebogenauswertungen eine Rolle. Um solche Mehrfachantworten kodieren und auswerten zu können, bietet SPSS zwei verschiedene Methoden an: die dichotome und die kategoriale Methode. Diese beiden Methoden werden in zwei getrennten Kapiteln jeweils anhand eines Beispiels dargestellt.

## 10.1 Dichotome Methode

Wir wollen die dichotome Methode am Beispiel einer Jugendbefragung erläutern. Eine Frage des Fragebogens lautete: »Bei welchen der folgenden Sachen machst du aktiv mit?« Es waren die folgenden Antwortmöglichkeiten vorgegeben:

- ▶ Freiwillige Schulveranstaltung
- ▶ Klassensprecher/SMV
- ▶ Natur-, Umwelt-, Tierschutz
- ▶ Menschenrechte
- ▶ Kirchengruppe
- ▶ Freiwillige Organisation
- ▶ Etwas selbstständig organisieren
- ▶ Sicherheit im Verkehr
- ▶ Gewerkschaft
- ▶ Parteien
- ▶ Stadtjugendring/Jugendverbände
- ▶ Politisch aktive Gruppe in der Stadt

Bei der Methode multipler Dichotomien wird für jede der Antwortmöglichkeiten eine eigene Variable definiert. Im gegebenen Beispiel werden dazu also zwölf Variablen benötigt. Kreuzt ein Jugendlicher die Antwort »Freiwillige Schulveranstaltung« an, wird der betreffenden Variablen der Code »1« zugeordnet, andernfalls eine »0«; kreuzt ein Jugendlicher die Antwort »Klassensprecher/SMV« an, wird der betreffenden Variablen der Code »1« zugeordnet, andernfalls eine »0« usw. So entstehen zwölf Variablen, jeweils mit 1 und 0 kodiert. Dabei ist die Wahl der Codezahlen natürlich beliebig; sie muss aber für alle Antwortmöglichkeiten gleich sein und dem Computer an der entsprechenden Stelle mitgeteilt werden.

Die Ergebnisse dieser Mehrfachfrage sind in der Datei mitmach.sav enthalten. Wir wollen zunächst eine Häufigkeitsauszählung der Frage »Bei welchen der folgenden Sachen machst du aktiv mit?« durchführen und dann eine Kreuztabelle dieser Frage mit dem Geschlecht erstellen.

### 10.1.1 Definition von Sets

Die Antworten auf unsere Frage sind in der beschriebenen Weise in den Variablen v1\_1 bis v1\_12 kodiert. In einem ersten Schritt müssen wir dem Computer mitteilen, dass diese zwölf Variablen zu einem »Variablenset« gehören.

- Laden Sie die Datei mitmach.sav.
- Treffen Sie die Menüwahl

*Analysieren*

*Mehrfachantworten*

*Variablensets definieren...*

Es öffnet sich die Dialogbox *Mehrfachantwortsets*.

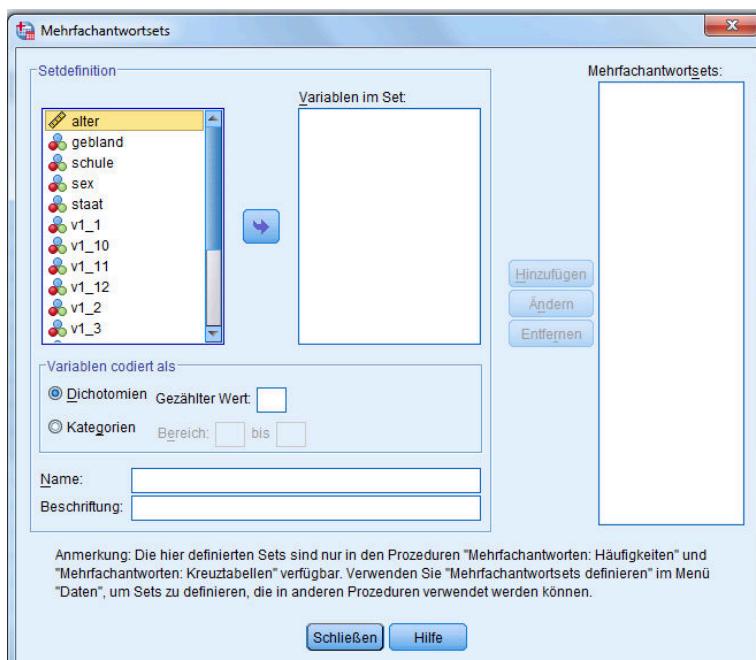


Bild 10.1: Dialogbox Mehrfachantwortsets

- Markieren Sie die Variablen v1\_1 bis v1\_12 in der Quellvariablenliste und übertragen Sie diese in die Zielvariablenliste *Variablen im Set*.
- Kennzeichnen Sie die Art der Kodierung als *Dichotomien* (dies ist die Voreinstellung). Legen Sie als zu zählenden Wert die »1« fest.

- Geben Sie dem Set den Namen »mitmach« und das Etikett »Mitarbeit der Jugendlichen«.
- Klicken Sie auf *Hinzufügen*, um das so generierte Set in die Liste der Mehrfachantworten-Sets aufzunehmen.

Setvariablen werden von SPSS mit einem führenden Dollarzeichen versehen; die erzeugte Setvariable trägt folglich den Namen \$mitmach.

- Klicken Sie auf *Schließen*, um den Vorgang der Set-Definition zu beenden.

Wir wollen nunmehr eine Häufigkeitsverteilung der soeben erzeugten Variablen erstellen.

### 10.1.2 Häufigkeitstabellen für dichotome Setvariablen

- Um Häufigkeitstabellen für dichotome Setvariablen zu erzeugen, wählen Sie aus dem Menü

*Analysieren*

*Mehrfachantworten*

*Häufigkeiten...*

Es öffnet sich die Dialogbox *Mehrfachantworten: Häufigkeiten*.

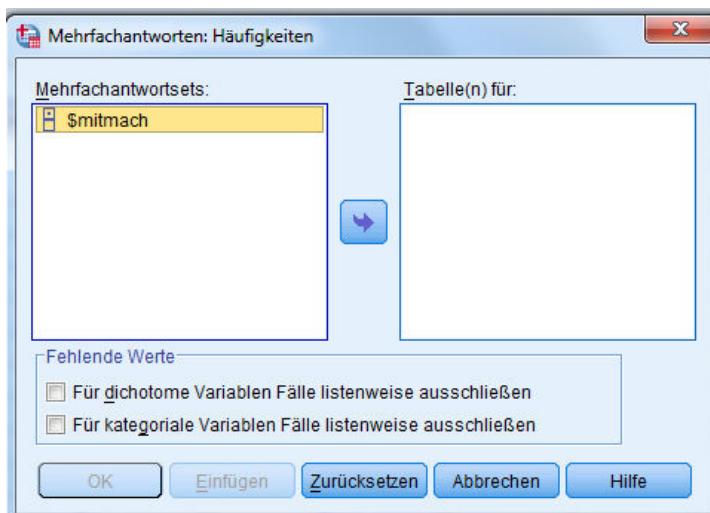


Bild 10.2: Dialogbox Mehrfachantworten: Häufigkeiten

Die aktuell definierten Mehrfachantworten-Sets erscheinen in der entsprechenden Liste, in unserem Beispiel ist es die Setvariable \$mitmach.

- Bringen Sie die Setvariable \$mitmach in die Liste *Tabelle(n) für*.
- Bestätigen Sie mit *OK*.

Im Viewer erscheint zunächst die Tabelle Fallzusammenfassung.

	Fallzusammenfassung					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
\$mitmach <sup>a</sup>	2936	61,6%	1830	38,4%	4766	100,0%

a. Dichotomie-Gruppe tabellarisch dargestellt bei Wert 1.

Die vorangestellte Tabelle *Fallzusammenfassung* liefert Informationen über die gültigen sowie die fehlenden Fälle. Von den 4.766 Fällen der Datendatei liegen 2.936 (61,6%) gültige Antworten sowie 1.830 (38,4%) fehlende Antworten vor. Als fehlend wird ein Fall genau dann bezeichnet, wenn keine der in die Setvariable einbezogenen Variablen einen gezählten Wert hat, d. h. im gegebenen Beispiel mit »1« kodiert ist.

Eine Variante erhalten Sie, wenn Sie in der Dialogbox *Mehrfachantworten: Häufigkeiten* die Option *Für dichotome Variablen Fälle listenweise ausschließen* aktivieren. Hier wird dann zusätzlich jeder Fall als fehlend deklariert, der in mindestens einer der in die Setvariable einbezogenen Variablen einen fehlenden Wert hat, in unserem Beispiel also weder mit 1 noch mit 0 kodiert wäre. Dies könnte z. B. dann relevant werden, wenn die betreffende Antwort im Fragebogen nicht eindeutig kenntlich gemacht ist.

Es folgt im Viewer die eigentliche Häufigkeitsverteilung:

		Antworten		Prozent der Fälle
		H	Prozent	
Mitarbeit der Jugendlichen <sup>a</sup>	Freiwillige Schulveranstaltung	1560	27,4%	53,1%
	Klassensprecher/SMV	917	16,1%	31,2%
	Natur-, Umwelt-, Tierschutz	600	10,5%	20,4%
	Menschenrechte	115	2,0%	3,9%
	Kirchengruppe	591	10,4%	20,1%
	Freiwillige Organisation	579	10,2%	19,7%
	Etwas selbstständig organisieren	1002	17,6%	34,1%
	Sicherheit im Verkehr	69	1,2%	2,4%
	Gewerkschaft	32	0,6%	1,1%
	Parteien	69	1,2%	2,4%
Gesamtsumme	Stadtjugendring/Jugendverbände	99	1,7%	3,4%
	Politisch aktive Gruppe in der Stadt	55	1,0%	1,9%
Gesamtsumme		5688	100,0%	193,7%

a. Dichotomiegruppe tabuliert bei Wert 1.

Zu den beobachteten Häufigkeiten werden zwei verschiedene Prozentuierungen angeboten: Zum einen wird auf die Gesamtzahl der gegebenen Ja-Antworten prozentuiert (5.688), zum anderen auf die Anzahl der gültigen Fälle (2.936). Die wohl griffigste Prozentuierung fehlt allerdings auch in der SPSS Version 20 noch immer: die Prozentuierung auf die Anzahl aller Fälle bzw. in unserem Fall aller Befragten der empirischen Erhebung (4.766).

Die erste Zeile der Häufigkeitstabelle können Sie z. B. wie folgt interpretieren: 1.560 befragte Jugendliche machen bei freiwilligen Schulveranstaltungen mit. Das sind 27,4% der gegebenen Antworten (5.688) und 53,1% derjenigen Befragten (2.936), die zumindest eine Antwort angekreuzt haben.

# Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwortschutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: [info@pearson.de](mailto:info@pearson.de)

## Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.**

## Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

**<http://ebooks.pearson.de>**