

Kategoriale Daten zusammenfassen: Häufigkeiten und Prozente



In diesem Kapitel ...

- ▶ Häufigkeitstabellen für kategoriale Daten erstellen
- ▶ Der kleine Unterschied zwischen absoluten Häufigkeiten und relativen Häufigkeiten
- ▶ Häufigkeitstabellen interpretieren und bewerten

Als *kategoriale Daten* werden solche Daten bezeichnet, die jede Beobachtung genau einer Kategorie zuordnen. Typische Beispiele für solche Kategorien sind klein, mittel, groß oder männlich, weiblich oder Deutscher, Franzose, Italiener, Österreicher etc. Wenn Sie kategoriale Daten erhoben haben, liegen diese in den meisten Fällen zunächst als lange Liste vor, in der jede befragte Person (beziehungsweise allgemein jede Beobachtungseinheit) eine Zeile bildet und die Werte der kategorialen Variablen in einer langen Spalte einzeln untereinander stehen. Eine solche unübersichtliche Liste ist jedoch nur selten aussagekräftig. Daher müssen die Daten für eine Auswertung zusammengefasst und in übersichtlicher Form dargestellt werden. Der einfachste Weg hierzu besteht darin auszuzählen, wie viele Beobachtungen in die einzelnen Kategorien fallen, und das Ergebnis als sogenannte *Häufigkeitstabelle* darzustellen. Genau dies können Sie in diesem Kapitel üben, nämlich das Erstellen und Interpretieren von absoluten und relativen Häufigkeitstabellen für kategoriale Daten.

Auf die Häufigkeit zählen

Die Anzahl der Beobachtungen, die in eine bestimmte Kategorie fallen, wird als *absolute Häufigkeit* dieser Kategorie bezeichnet. Durch das einfache Auflisten sämtlicher Kategorien mit den dazugehörigen Häufigkeiten erhalten Sie eine Häufigkeitstabelle. Die Summe der Häufigkeiten aller Kategorien sollte dabei stets der Größe der gesamten Stichprobe entsprechen, da ja schließlich jede Beobachtung genau einer Kategorie zugeordnet wurde.

Die folgenden Aufgaben enthalten Beispiele dafür, wie sich kategoriale Daten mithilfe von Häufigkeitstabellen zusammenfassen lassen.



Angenommen, Sie haben zehn Personen gefragt, ob sie ein Handy besitzen. Jede dieser zehn Personen kann also in eine von zwei möglichen Kategorien fallen: *Ja* oder *Nein*. Das Ergebnis Ihrer Befragung ist in der folgenden Tabelle wiedergegeben.

Person Nr.	Handy	Person Nr.	Handy
1	Ja	6	Ja
2	Nein	7	Ja
3	Ja	8	Ja
4	Nein	9	Nein
5	Ja	10	Ja

1. Fassen Sie diese Daten in einer Häufigkeitstabelle zusammen.
2. Worin besteht der Nutzen der Zusammenfassung von kategorialen Daten?

Lösung

In der Zusammenfassung werden die erhobenen Daten klar und übersichtlich dargestellt.

1. Die Häufigkeitstabelle für die Antworten der zehn befragten Personen ist in der folgenden Tabelle wiedergegeben.
2. Die Zusammenfassung der Daten ermöglicht es Ihnen, die Relevanz der einzelnen Kategorien abzulesen und Muster in den Daten zu erkennen, die aus Rohdaten nicht auf den ersten Blick ersichtlich wären.

Besitzen Sie ein Handy?	Absolute Häufigkeit
Ja	7
Nein	3
Gesamt	10

Aufgabe 1

Sie lassen 20 Kunden in einem Supermarkt zwei unterschiedliche Marken eines Softdrinks testen und fragen anschließend, welches Getränk besser geschmeckt hat, Marke A oder Marke B. Die Ergebnisse dieser Befragung sehen wie folgt aus: A, A, B, B, B, B, B, B, A, A, A, B, A, A, A, A, B, B, A, A.

Welche der beiden Marken bevorzugen die Kunden? Erstellen Sie hierzu eine Häufigkeitstabelle und erläutern Sie das Ergebnis.

Aufgabe 2

Die Einwohner einer Stadt sind in einer Volksabstimmung dazu aufgerufen, über die Erhöhung einer Sonderabgabe zur Finanzierung der öffentlichen Schulen abzustimmen. Insgesamt haben 18.726 Wähler ihre Stimme hierzu abgegeben. Davon stimmten 10.479 für die Erhöhung der Sonderabgabe, die übrigen Wähler dagegen.

1. Stellen Sie das Ergebnis in einer Häufigkeitstabelle dar.
2. Warum ist es wichtig, die Anzahl der insgesamt befragten Personen in der Tabelle mit auszuweisen?

Aufgabe 3

Ein Zoo führt eine Befragung durch und möchte von 1.000 Personen wissen, ob sie im letzten Jahr den Zoo besucht haben. 592 der Befragten antworten mit Ja, 198 mit Nein und 210 haben gar nicht geantwortet.

1. Erstellen Sie eine Häufigkeitstabelle mit den Ergebnissen der Befragung.
2. Erläutern Sie, warum es notwendig ist, auch die Personen, die keine Antwort abgegeben haben, in der Häufigkeitstabelle mit aufzuführen.

Aufgabe 4

Mal angenommen, Sie würden eine Häufigkeitstabelle erstellen, in der für jede Kategorie nicht die absolute Anzahl der Nennungen dieser Kategorie aufgeführt wird, sondern lediglich der Prozentwert (die *relative Häufigkeit*). Welchen Vorteil hat eine solche relative Häufigkeitstabelle gegenüber einer Häufigkeitstabelle mit absoluten Häufigkeiten?

Kategorien vergleichen mit Prozentwerten

Eine andere Möglichkeit, kategoriale Daten zusammenzufassen und in verdichteter Form darzustellen, besteht darin aufzulisten, welcher prozentuale Anteil der Antworten auf die verschiedenen Kategorien entfällt. Diese Prozentwerte sind die *relativen Häufigkeiten* der Kategorien. Die *relative Häufigkeit* einer bestimmten Kategorie ergibt sich aus der absoluten Häufigkeit (der Anzahl der Beobachtungen in einer Kategorie) geteilt durch die gesamte Stichprobengröße. Wenn Sie beispielsweise 50 Personen nach ihren Präferenzen zu einem bestimmten Thema befragen und zehn der Befragten geben sich als Anhänger einer bestimmten Variante zu erkennen, errechnet sich die relative Häufigkeit der Anhänger dieser Variante als $10 \div 50 = 0,2$ beziehungsweise 20 %.

Durch die Auflistung sämtlicher Kategorien mit ihren jeweiligen relativen Häufigkeiten erstellen Sie eine *relative Häufigkeitstabelle*. Die Summe der relativen Häufigkeiten aller Kategorien sollte stets 100 Prozent ergeben (wenn man einmal kleine Rundungsfehler außer Acht lässt).

Das folgende Beispiel zeigt, wie kategoriale Daten in einer relativen Häufigkeitstabelle zusammengefasst werden können.



In der folgenden Tabelle können Sie für insgesamt zehn Personen ablesen, ob diese ein Mobiltelefon besitzen. Erstellen Sie für diese Daten eine relative Häufigkeitstabelle und interpretieren Sie das Ergebnis.

Person Nr.	Handy	Person Nr.	Handy
1	Ja	6	Ja
2	Nein	7	Ja
3	Ja	8	Ja
4	Nein	9	Nein
5	Ja	10	Ja

Lösung

Die folgende Tabelle zeigt eine relative Häufigkeitstabelle für die Handy-Daten. 70 % der befragten Personen haben angegeben, ein Handy zu besitzen, während 30 % offenbar noch in der Steinzeit leben und mit Rauchzeichen kommunizieren.

Besitzen Sie ein Handy?	Relative Häufigkeit
Ja	70 %
Nein	30 %

Den Wert 70 % in der Tabelle errechnen Sie, indem Sie die Zahl der Personen mit Handy (also 7) durch die Zahl aller befragten Personen (10) dividieren. Sie rechnen also $7 \div 10 = 0,7$ beziehungsweise 70 %. Entsprechend ergeben sich die 30 % »Steinzeitmenschen« als $3 \div 10 = 0,3$ beziehungsweise 30 %.

Aufgabe 5

Sie lassen 20 Kunden in einem Supermarkt zwei verschiedene Softdrinks testen und fragen anschließend, welches Getränk ihnen besser schmeckt, Marke A oder Marke B. Sie erhalten die folgenden Ergebnisse: A, A, B, B, B, B, B, A, A, A, B, A, A, A, A, B, B, A, A. Welches Getränk findet bei den Befragten mehr Zuspruch?

1. Erstellen Sie eine relative Häufigkeitstabelle, um das beliebteste Getränk zu ermitteln.
2. Welche Art von Häufigkeitsauswertung ist generell leichter zu interpretieren: absolute oder relative Häufigkeiten? Begründen Sie Ihre Antwort.

Aufgabe 6

Die Bewohner einer Stadt sind aufgerufen, in einer Volksabstimmung über die Erhöhung einer Sonderabgabe zur Finanzierung öffentlicher Schulen zu entscheiden. Insgesamt haben 18.726 Wähler ihre Stimme hierzu abgegeben; davon stimmten 10.479 für die Erhöhung der Sonderabgabe, die übrigen Wähler dagegen. Stellen Sie das Wahlergebnis in einer relativen Häufigkeitstabelle dar.

Aufgabe 7

Ein Zoo führt eine Befragung durch und möchte von 1.000 Personen wissen, ob sie im letzten Jahr den Zoo besucht haben. 592 der Befragten antworten mit Ja, 198 mit Nein und 210 haben gar nicht geantwortet. Erstellen Sie für dieses Ergebnis eine relative Häufigkeitstabelle und ermitteln Sie anhand der Tabelle die Antwortrate, also den Anteil der Personen, die überhaupt auf die Frage nach dem Zoobesuch geantwortet haben.

Aufgabe 8

Nennen Sie einen Nachteil von relativen Häufigkeitstabellen gegenüber einfachen Häufigkeitstabellen mit absoluten Häufigkeiten.

Vorsicht bei der Interpretation von absoluten und relativen Häufigkeiten

Werden umfangreiche kategoriale Daten wie in einer Häufigkeitstabelle zu verdichteten Daten zusammengefasst, können dabei leicht wichtige Informationen verloren gehen, die in den Ursprungsdaten enthalten sind. Zusammengefasste Daten bergen daher stets das Risiko, dass sie unpräzise oder unvollständig sind. Bei der Interpretation von Häufigkeitstabellen und anderen Formen von Datenzusammenfassungen sollten Sie daher stets wissen, auf welche möglichen Schwachstellen Sie achten müssen, um nicht irreführende oder unvollständige Informationen zu erhalten.



In Statistik-Kursen und -Prüfungen fordern die Lehrer gerne dazu auf, »die Ergebnisse zu interpretieren«. Was der Prüfer damit eigentlich meint, ist nichts anderes, als die vorliegenden Daten in Bezug auf die jeweilige Fragestellung zu bewerten. Mit anderen Worten will der Prüfer wissen: Was kann die Person, die die vorliegenden Daten vor dem Hintergrund einer konkreten Fragestellung erhoben hat, nun aus den gewonnenen Daten lernen?



Wenn Sie relative Häufigkeitstabellen erstellen oder auswerten, vergessen Sie nie zu prüfen, ob sich die Prozentwerte sämtlicher Kategorien zu 100 % addieren (abgesehen von minimalen Rundungsfehlern, die zulässig sind). Außerdem sollten Sie sich nie auf die alleinige Betrachtung der relativen Häufigkeiten beschränken, sondern stets überprüfen, wie groß die Stichprobe ist, auf der die relative Häufigkeitstabelle basiert.

Das folgende Beispiel zeigt einige entscheidende Aspekte für einen kritischen Umgang mit zusammengefassten Daten.



Sie sehen im Fernsehen einen Werbespot, in dem der Hersteller eines neuen Erkältungsmedikaments »Antigrippe« dieses mit dem derzeit meistverkauften Erkältungsmedikament vergleicht. Die Ergebnisse sind in der folgenden Tabelle wiedergegeben.

Bewertung von Antigrippe im Vergleich zum Marktführer	Prozent
Viel besser	47 %
Mindestens gleich gut	18 %

1. Was für eine Art von Tabelle wurde hier verwendet?
2. Interpretieren Sie das Ergebnis. Hat das neue Erkältungsmedikament den aktuellen Marktführer geschlagen?
3. Welche wichtigen Informationen fehlen in der Tabelle?

Lösung

Mit der im Werbespot verwendeten Tabelle verhält es sich ähnlich wie mit guten Ratschlägen: Sie hilft nicht wirklich weiter.

1. Bei der Tabelle handelt es sich um eine unvollständige relative Häufigkeitstabelle. Die Tabelle ist unvollständig, weil die Kategorie »Antigrippe ist *weniger gut* als das führende Produkt« fehlt. Sie können den fehlenden Wert aber leicht selbst berechnen, denn vermutlich sagen $100\% - (47\% + 18\%) = 35\%$ der Testpersonen, dass der aktuelle Marktführer besser ist als Antigrippe. (Wenn es nicht noch weitere Kategorien wie »Weiß nicht« gibt.)
2. Wenn man die beiden Kategorien aus der Tabelle zusammenfasst, erkennt man, dass 65 % der Testpersonen angeben, dass Antigrippe mindestens so gut wirkt wie das aktuell führende Produkt. Etwa die Hälfte der Personen sagt sogar, Antigrippe wirke besser.
3. Welche Informationen fehlen? Zunächst einmal fehlt mindestens eine Kategorie, damit sich alle möglichen Antworten fair miteinander vergleichen lassen. Noch gravierender ist aber, dass die Stichprobengröße nicht mit ausgewiesen wird. So wird nicht deutlich, ob die Ergebnisse auf einer Befragung von 10, 100 oder 1.000 Patienten basieren. Damit ist die Zuverlässigkeit der Ergebnisse vollkommen unklar. Es ist überhaupt nicht zu erkennen, ob sich die Ergebnisse verallgemeinern lassen (weil die Stichprobe so groß war, dass starke Abweichungen vom tatsächlichen Stimmungsbild sehr unwahrscheinlich sind) oder ob die Ergebnisse letztlich nur für einige wenige von dem Hersteller befragte Patienten gelten.

Aufgabe 9

Sie haben 1.000 Personen gebeten, aus einer Liste von typischen Urlaubszielen diejenigen anzugeben, die der jeweilige Befragte bereits besucht hat. Bei dieser Befragung haben Sie folgende Ergebnisse erhalten: Spanien: 216 Nennungen; Italien: 312; Frankreich: 418; England: 359; USA: 188.

1. Erläutern Sie, warum eine klassische relative Häufigkeitstabelle für diese Daten nicht sinnvoll ist.
2. Wie können Sie die vorliegenden Daten stattdessen sinnvoll mithilfe von relativen Häufigkeiten zusammenfassen und darstellen?

Aufgabe 10

Angenommen, Ihnen liegt als Ergebnis einer Befragung ausschließlich eine Häufigkeitstabelle mit absoluten Häufigkeiten vor. Können Sie anhand dieser Tabelle die korrespondierende relative Häufigkeitstabelle erstellen?

Wie verhält es sich im umgekehrten Fall? Können Sie auf Basis einer relativen Häufigkeitstabelle auch die entsprechende Tabelle mit absoluten Häufigkeiten erstellen?

Begründen Sie Ihre Antworten.

Lösungen für die Aufgaben zum Thema Zusammenfassen von kategorialen Daten

Lösung zur Aufgabe 1.1

11 Kunden bevorzugen Marke A, 9 Kunden geben Marke B den Vorzug. Dieses Ergebnis ist in der folgenden Häufigkeitstabelle dargestellt. Marke A hat damit mehr Stimmen erhalten, insgesamt liegen beiden Marken jedoch sehr dicht beieinander.

Bevorzugte Marke	Absolute Häufigkeit
Marke A	11
Marke B	9
Gesamt	20

Lösung zur Aufgabe 1.2

Kategoriale Daten können sehr gut durch die einfache Angabe der absoluten Häufigkeiten der einzelnen Kategorien zusammengefasst werden, solange man dabei stets die Gesamtzahl aller Beobachtungen im Auge behält.

- Die Ergebnisse der Volksabstimmung sind in der folgenden Tabelle dargestellt. Da insgesamt 18.726 Einwohner abgestimmt haben und davon 10.479 mit Ja stimmten, ergibt sich die Anzahl der Nein-Stimmen aus der Differenz und beträgt damit $18.726 - 10.479 = 8.247$.
- Es ist wichtig, die Gesamtzahl der Befragten zu kennen, da sich nur im Vergleich zu dieser Gesamtzahl die Häufigkeiten einzelner Kategorien bewerten lassen.

Wahl	Absolute Häufigkeit
Ja	10.479
Nein	8.247
Gesamt	18.726

Lösung zur Aufgabe 1.3

Dieses Beispiel zeigt, wie wichtig es ist, bei der Ergebnisdarstellung nicht nur die Antworten der Befragungsteilnehmer auszuweisen, sondern auch anzugeben, wie viele Personen insgesamt gefragt wurden und welcher Anteil davon tatsächlich an der Befragung teilgenommen hat.

- Die Ergebnisse der Befragung werden in der folgenden Häufigkeitstabelle wiedergegeben.
- Würden in dieser Tabelle die Befragten ohne Antwort nicht mit ausgewiesen werden, würden sich die Häufigkeiten der einzelnen Kategorien nicht zu 1.000 (der Gesamtzahl der Befragten) addieren. Grundsätzlich besteht natürlich die Möglichkeit, die Ergebnisse nur auf Basis der gültigen Antworten darzustellen, diese Darstellung könnte jedoch verzerrt sein und relevante Informationen verschweigen. Würde man die Befragten ohne Antwort

einfach unter den Tisch fallen lassen, würde man damit implizit unterstellen, diese Personen hätten, wenn sie denn an der Befragung teilgenommen hätten, im Wesentlichen genauso geantwortet wie die tatsächlichen Befragungsteilnehmer. Diese Annahme ist jedoch nicht ohne Weiteres zulässig.

Haben Sie im letzten Jahr den Zoo besucht?	Absolute Häufigkeit
Ja	592
Nein	198
keine Antwort	210
Gesamt	1.000

Lösung zur Aufgabe 1.4

Wenn Sie für die Zusammenfassung kategorialer Daten statt absoluter Häufigkeiten Prozentwerte angeben, erzeugen Sie damit eine relative Häufigkeitstabelle. Ein Vorteil einer solchen relativen Häufigkeitstabelle besteht darin, dass sich die Anteile der einzelnen Kategorien (also ihre relativen Häufigkeiten) stets zu 100 % addieren. Dadurch fällt es häufig leichter, die Ergebnisse zu interpretieren, da sich die Bedeutung einzelner Kategorien in Relation zu den übrigen Kategorien leichter ablesen lässt, insbesondere wenn viele verschiedene Kategorien betrachtet werden. Ferner können verschiedene Tabellen gleichen Aufbaus miteinander verglichen werden.

Lösung zur Aufgabe 1.5

Relative Häufigkeiten leisten genau das, was ihr Name verspricht: Sie helfen dabei, die einzelnen Ergebnisse in Relation zueinander zu setzen.

- 11 von 20 Kunden bevorzugen Marke A, 9 der 20 Befragten haben Marke B gewählt. Die relative Häufigkeitstabelle für dieses Ergebnis ist in der folgenden Darstellung wiedergegeben. Marke A hat mehr Stimmen erhalten, allerdings liegen beide Marken nahe beieinander: 55 % der Kunden präferieren Marke A, 45 % geben Marke B den Vorzug.

Bevorzugte Marke	Relative Häufigkeit
Marke A	55 %
Marke B	45 %

2. Häufig lassen sich Prozentangaben leichter interpretieren als absolute Häufigkeiten, denn für eine sinnvolle Bewertung von absoluten Häufigkeiten müssen diese immer noch in Relation zur Basis gesetzt werden, Sie benötigen also Aussagen der Art »soundso viele von insgesamt soundso vielen«.

Lösung zur Aufgabe 1.6

Die Ergebnisse sind in der folgenden relativen Häufigkeitstabelle wiedergegeben. Der Prozentwert der Ja-Stimmen beträgt $10.479 \div 18.726 = 55,96\%$. Da die Gesamtheit stets 100 % umfasst, lässt sich der Anteil der Nein-Stimmen berechnen als $100\% - 55,96\% = 44,04\%$.

Wahl	Relative Häufigkeit
Ja	55,96 %
Nein	44,04 %

Lösung zur Aufgabe 1.7

Die folgende Tabelle gibt die relative Häufigkeitstabelle für die Befragungsergebnisse wieder. Um die Ergebnisse angemessen interpretieren zu können, ist es wichtig, auch die Antwortrate zu kennen. Je höher die Antwortrate, desto besser. Die Antwortrate beträgt hier $59,2\% + 19,8\% = 79\%$. Dies ist der Gesamtanteil der Personen, die auf die Frage in irgendeiner Form (mit Ja oder Nein) geantwortet haben. 21 Prozent haben dagegen nicht geantwortet.

Haben Sie im letzten Jahr den Zoo besucht?	Relative Häufigkeit
Ja	$592 \div 1.000 = 0,592 = 59,2\%$
Nein	19,8 %
keine Antwort	21 %

Lösung zur Aufgabe 1.8

Ein Nachteil einer relativen Häufigkeitstabelle besteht darin, dass sie nur die prozentuale Verteilung der Ergebnisse anzeigt, nicht aber die absolute Anzahl der Beobachtungen, die der Tabelle zugrunde liegen. Ohne diese Information lässt sich jedoch nicht einschätzen, wie zuverlässig die Ergebnisse sind und ob ihre Aussagekraft hinreichend groß ist, um die Ergebnisse zu verallgemeinern. Diese Schwäche von relativen Häufigkeitstabellen können Sie vermeiden, indem Sie einfach die Größe der Stichprobe direkt ober- oder unterhalb der Tabelle mit angeben.



Wenn Sie eine relative Häufigkeitstabelle erstellen, geben Sie gemeinsam mit der Tabelle stets die Stichprobengröße an.

Lösung zur Aufgabe 1.9

Besonders vorsichtig müssen Sie bei der Interpretation von solchen Tabellen sein, bei denen es möglich ist, dass eine Person (beziehungsweise allgemein eine Beobachtung) mehreren Kategorien gleichzeitig zugeordnet ist.

- Die Häufigkeiten der einzelnen Kategorien addieren sich hier nicht zu 1.000, der Stichprobengröße, da jeder Befragte eines, mehrere oder auch gar keines der Reiseziele aus der Liste auswählen kann. Es ist also nicht sichergestellt, dass jeder Befragte im Ergebnis genau einer Kategorie zugeordnet ist. Wenn Sie die Summe aller Antworten berechnen (1.493) und die Häufigkeit jeder einzelnen Kategorie durch 1.493 teilen, erhalten Sie die relative Häufigkeit dieser Kategorie. Die relativen Häufigkeiten sämtlicher Kategorien addieren sich dabei auch zu 100 %, genau so, wie es sein soll. Aber was sagen die relativen Häufigkeiten aus? Es ist sehr schwierig, diese Prozentwerte zu interpretieren, denn sie basieren nicht auf der Anzahl der befragten Personen (sondern auf der Anzahl der insgesamt genannten Reiseziele).

2. Eine Möglichkeit, die Ergebnisse sinnvoll darzustellen, besteht darin, für jedes Urlaubsziel den Anteil der Personen anzugeben, die dieses Urlaubsziel bereits besucht haben (im Vergleich zu den Personen, die das Urlaubsziel noch nicht kennen). Diese Prozentangaben beziehen sich nur jeweils auf das einzelne Urlaubsziel und addieren sich daher für jedes einzelne Urlaubsziel zu 100 % (Anteil der Besucher + Anteil der Nichtbesucher = 100 %). Die folgende Tabelle stellt die Ergebnisse in dieser Form dar. *Beachten Sie aber:* Bei dieser Tabelle handelt es sich nicht um eine relative Häufigkeitstabelle, auch wenn die Darstellung freilich relative Häufigkeiten nutzt.

Urlaubsziel	% Besucher	% Nichtbesucher
Spanien	$216 \div 1.000 = 21,6 \%$	$100 \% - 21,6 \% = 78,4 \%$
Italien	$312 \div 1.000 = 31,2 \%$	68,8 %
Frankreich	$418 \div 1.000 = 41,8 \%$	58,2 %
England	$359 \div 1.000 = 35,9 \%$	64,1 %
USA	$188 \div 1.000 = 18,8 \%$	81,2 %



Nicht in allen Tabellen, die Prozentangaben ausweisen, müssen sich diese zu 1 (beziehungsweise 100 %) addieren. Verbiegen Sie eine Tabelle daher nicht so lange, bis sich eine Summe von 100 % ergibt, wenn dies gar nicht erforderlich ist. Hinterfragen Sie besser stets, ob es sein kann, dass eine Person (oder allgemein eine Beobachtung) in mehrere Kategorien fällt. Ist dies der Fall, ist eine relative Häufigkeitstabelle, in der sich die einzelnen relativen Häufigkeiten zu 100 % addieren, für die Darstellung der Ergebnisse nicht geeignet.

Lösung zur Aufgabe 1.10

Sie können immer einfach sämtliche absoluten Häufigkeiten zu einer Gesamthäufigkeit addieren und anschließend die relativen Häufigkeiten berechnen, indem Sie die einzelnen absoluten Häufigkeiten durch die Gesamthäufigkeit teilen.

Liegen Ihnen dagegen nur die Prozentangaben vor, können Sie nicht ohne Weiteres rückwärts vorgehen und die ursprünglichen absoluten Häufigkeiten ermitteln, es sei denn, Ihnen ist auch die Anzahl der Personen (Beobachtungen) bekannt, die insgesamt befragt wurden. Angenommen, Sie wissen, dass 80 % der Personen aus einer Umfrage Eiscreme mögen. Wie viele Personen mögen dann Eiscreme? Wenn insgesamt 100 Personen befragt wurden, mögen 80 Befragte ($100 \cdot 0,8$) Eiscreme. Umfasst die Stichprobe dagegen nur 50 Befragte, mögen 40 der Befragten ($50 \cdot 0,8$) Eiscreme. Wurden insgesamt nur 5 Personen befragt, mögen nur 4 davon ($5 \cdot 0,8$) Eiscreme. Diese Rechnungen zeigen noch einmal, wie wichtig es ist, dass eine relative Häufigkeitstabelle an irgendeiner Stelle auch die Größe der Stichprobe mit ausweist.



Achten Sie bei der Auswertung einer relativen Häufigkeitstabelle stets auf die Größe der Stichprobe. Lassen Sie sich nicht von einer isolierten Prozentangabe in die Irre führen, indem Sie implizit unterstellen, diese Angaben würden schon auf einer großen Anzahl von Beobachtungen basieren. Häufig ist dies nicht der Fall.