

---

# 2. Eindimensionale Häufigkeitsverteilungen

---

Thema dieses Abschnitts ist die Auswertung eindimensionalen (univariaten) Datenmaterials, d.h. Daten zu einem einzigen Merkmal einer Grundgesamtheit oder Stichprobe. Der erste Schritt besteht dabei im Auszählen von Häufigkeiten, mit der die Ausprägungen des interessierenden Merkmals einzeln oder in Klassen zusammengefasst auftreten. Dies liefert uns die sog. Häufigkeitsverteilung. In einem zweiten Schritt können wir Maßzahlen berechnen, die diese Häufigkeitsverteilung charakterisieren. Maßzahlen geben zwar einen kompakten Überblick über wesentliche Eigenschaften einer Häufigkeitsverteilung, doch lassen sich in grafischen Darstellungen oft Strukturen im Datenmaterial erkennen, die Maßzahlen nicht widerspiegeln. Gerade daher empfiehlt es sich, tabellarische und grafische Darstellungen von Häufigkeitsverteilungen mit ihren Maßzahlen zu kombinieren, um das vorliegende Datenmaterial möglichst gut zu beschreiben.

## 2.1 Häufigkeitsverteilung

Das Auszählen der Häufigkeiten von Ausprägungen diskreter Merkmale und derer stetiger (bzw. quasistetiger) Merkmale erfordert unterschiedliche Vorgehensweisen. Bei diskreten Merkmalen (mit nicht zu vielen unterschiedlichen möglichen Ausprägungen) kommen im auszuwertenden Datenmaterial die einzelnen Ausprägungen meist mehrmals, d.h. bei mehreren Merkmalsträgern vor. Bei stetigen (und quasistetigen) Merkmalen ist das mehrmalige Auftreten hingegen nicht die Regel.

### 2.1.1 Häufigkeitsverteilung bei diskreten Merkmalen

In statistischen Erhebungen werden von einer Grundgesamtheit mit insgesamt  $N$  Elementen (Merkmalsträgern) in der Regel genau  $n$  Elemente (Stichprobe) hinsichtlich eines einzigen Merkmals  $X$  untersucht. Die durch die Stichprobe gewonnenen Werte  $x_1, x_2, \dots, x_n$  bezeichnen wir als Beobachtungswerte. Durch Aneinanderreihung aller Beobachtungswerte erhalten wir die sog. Beobachtungsreihe oder **Urliste**. Anders ausgedrückt ist die Urliste das  $n$ -Tupel  $(x_1, x_2, \dots, x_n)$  aller  $n$  Beobachtungswerte.

Zur Aufbereitung dieses Datenbestandes bzw. der Urliste empfiehlt es sich, die Häufigkeitsverteilung des Merkmals  $X$  zu gewinnen. Dazu benötigen wir die absoluten und relativen Häufigkeiten, der in der Urliste vorkommenden individuellen Merkmalsausprägungen von  $X$ . Wir bezeichnen diese Merkmalsausprägungen mit  $a_1, a_2, \dots, a_k$ , d.h. wir gehen davon aus, dass in der Stichprobe genau  $k$  verschiedene Ausprägungen des Merkmals  $X$  auftreten und  $k$  nicht zu groß ist. Wir können damit für die gesuchten Häufigkeiten die folgenden Zusammenhänge festhalten:

## 1. Absolute Häufigkeit

Unter der absoluten Häufigkeit  $h(a_j)$  verstehen wir die Anzahl von Elementen der Urliste, die auf eine bestimmte Merkmalsausprägung  $a_j$  entfallen.

Es gilt allgemein

$$0 \leq h(a_j) \leq n \quad (I2: 1)$$

und

$$h(a_1) + h(a_2) + \dots + h(a_k) = \sum_{j=1}^k h(a_j) = n. \quad (I2: 2)$$

Der Wert der absoluten Häufigkeit muss also nach (I2: 1) zwingend größer oder gleich Null und kleiner oder gleich der Gesamtzahl  $n$  der Elemente der Urliste sein. Zudem ergibt nach (I2: 2) die Summe aller absoluten Häufigkeiten stets  $n$ .

## 2. Relative Häufigkeit

Dividieren wir die absoluten Häufigkeiten  $h(a_j)$  durch die Gesamtzahl der Elemente  $n$ , so erhalten wir die relative Häufigkeit

$$f(a_j) = \frac{h(a_j)}{n}, \quad (I2: 3)$$

für die

$$0 \leq f(a_j) \leq 1 \quad (I2: 4)$$

und

$$f(a_1) + f(a_2) + \dots + f(a_k) = \sum_{j=1}^k f(a_j) = 1 \quad (I2: 5)$$

gilt. Da die relative Häufigkeit  $f(a_j)$  nichts anderes angibt als den Anteil, mit dem  $a_j$  in der Stichprobe des Umfangs  $n$  vertreten ist, liegt  $f(a_j)$  nach (I2: 4) logischerweise stets zwischen 0 und 1 oder stimmt mit diesen Werten überein. Zudem muss die Summe aller Anteilswerte zwingend 1 ergeben, was (I2: 5) zum Ausdruck bringt. Relative Häufigkeiten werden häufig auch in Prozent angegeben. Ihre Summe liegt dann bei 100 %.

Unter der **absoluten Häufigkeitsverteilung** verstehen wir nun nichts anderes als die Paare  $(a_j, h(a_j))$  mit  $j = 1, 2, \dots, k$  und unter der **relativen Häufigkeitsverteilung** die Paare  $(a_j, f(a_j))$  mit  $j = 1, 2, \dots, k$ . Die einzelnen Merkmalsausprägungen  $a_j$  werden in einer Häufigkeitsverteilung meist zur besseren Übersicht in aufsteigender Reihenfolge sortiert.

Häufigkeitsverteilungen können wir übersichtlich in Form von *Häufigkeitstabellen* darstellen. Es empfiehlt sich dabei die Verwendung der in Abbildung I 5 skizzierten Struktur. Wir erkennen daran recht schön, dass die Eigenschaften (I2: 2) und (I2: 5) von absoluten und relativen Häufigkeiten hier als Spaltensummen überprüfbar sind. Für eine anschauliche Tabellengestaltung empfiehlt es sich, mit möglichst wenig Linierungen zu arbeiten, um die Tabelle optisch nicht zu überfrachten.

Laufende Nummer	Merkmalsausprägung	absolute Häufigkeit	relative Häufigkeit
j	$a_j$	$h(a_j)$	$f(a_j)$
1	$a_1$	$h(a_1)$	$f(a_1)$
2	$a_2$	$h(a_2)$	$f(a_2)$
...	...	...	...
k	$a_k$	$h(a_k)$	$f(a_k)$
$\Sigma$	-	n	1

Abbildung I 5: Häufigkeitstabelle

**Beispiel I 2-1**

Tabellarische Darstellung einer Häufigkeitsverteilung 

Anlässlich einer Mitarbeiterstatistik wurde in einer Abteilung eines Unternehmens das Alter X von n = 25 Mitarbeitern erfasst. Dabei ergab sich die folgende Urliste:

i	$x_i$ in Jahren	i	$x_i$ in Jahren	i	$x_i$ in Jahren
1	23	10	23	19	35
2	25	11	25	20	25
3	23	12	29	21	23
4	29	13	25	22	35
5	35	14	29	23	25
6	25	15	25	24	29
7	23	16	23	25	25
8	25	17	35		
9	29	18	25		

Wir erkennen daran, dass in der Stichprobe lediglich die k = 4 Ausprägungen 23, 25, 29 und 35 des Merkmals X auftauchen. Durch Auszählung dieser Ausprägungen erhalten wir die absoluten Häufigkeiten. Diese liefern dividiert durch n = 25 schließlich die relativen Häufigkeiten. Wir können damit die Häufigkeitsverteilung tabellarisch wie folgt darstellen:

Laufende Nummer	Merkmalsausprägung	Absolute Häufigkeit	Relative Häufigkeit	Prozentwert
j	$a_j$	$h(a_j)$	$f(a_j)$	100 % · $f(a_j)$
1	23	6	0,24	24 %
2	25	10	0,40	40 %
3	29	5	0,20	20 %
4	35	4	0,16	16 %
$\Sigma$	-	25	1	100 %

Alternativ zur Häufigkeitstabelle kann die Darstellung der absoluten und relativen Häufigkeitsverteilung auch *grafisch* in Form eines Piktogrammes, Stab- oder Säulendiagrammes, Häufigkeitspolygons oder Kreisdiagrammes erfolgen. Auch wenn sich in der Praxis überwiegend das Säulendiagramm durchgesetzt hat, wollen wir im Folgenden dennoch alle Möglichkeiten kurz betrachten:

## 1. Piktogramm

Wie die Diagrammbezeichnung schon verrät, erfolgt eine Darstellung der Häufigkeiten durch unterschiedliche Anzahl oder Größe von Bildsymbolen.

## 2. Stabdiagramm

Ein Stabdiagramm (höhenproportionale Darstellung) entsteht durch Eintragung von Punkten  $P(a_j, h(a_j))$  oder  $P(a_j, f(a_j))$  in ein Koordinatensystem, wobei jeder Punkt durch eine Senkrechte mit der x-Achse bzw.  $a_j$ -Achse verbunden wird. Die Längen dieser Senkrechten (Stäbe) entsprechen daher den absoluten bzw. relativen Häufigkeiten. Werden relative Häufigkeiten abgetragen, so muss die Gesamtlänge aller Stäbe zusammen gleich 1 sein. Wir können daher bei Stabdiagrammen für relative Häufigkeiten immer den gleichen Maßstab auf der y-Achse wählen. Dies bietet sich insbesondere beim Vergleich der Ergebnisse mehrerer Stichproben an.

## 3. Häufigkeitspolygon

Wie bei der Konstruktion eines Stabdiagrammes werden hier zunächst entweder die Punkte  $P(a_j, h(a_j))$  oder  $P(a_j, f(a_j))$  in ein Koordinatensystem eingetragen. Im Anschluss daran werden diese durch Geraden verbunden.

## 4. Säulendiagramm

Im Vergleich zum Stabdiagramm werden beim Säulendiagramm (flächenproportionale Darstellung) die Häufigkeiten durch Flächen (Rechtecke) und nicht durch Strecken abgebildet. Um eine Verwechslung mit einem Histogramm (vgl. Abschnitt I 2.1.3) zu vermeiden, ist zu beachten, dass die Rechtecke nicht aneinanderstoßen.

## 5. Kreissektorendiagramm

Hier wird zu jeder Merkmalsausprägung  $a_j$  ein Kreissektor gebildet, wobei die Flächen der Sektoren und damit auch die Innenwinkel proportional zu den Häufigkeiten gewählt werden. Dazu muss für den Innenwinkel  $\alpha_j$  des j-ten Sektors

$$\alpha_j = 360^\circ \cdot f(a_j) \quad (I2: 6)$$

gelten. Kreissektorendiagramme eignen sich besonders gut für nominalskalierte Merkmale, da hier die Merkmalsausprägungen nicht auf einem Zahlenstrahl abgetragen werden müssen und somit aufgrund ihrer Anordnung nicht fälschlicherweise eine Rangfolge vermutet wird.



### Beispiel I 2-2

### Grafische Darstellung einer Häufigkeitsverteilung

Eine Untersuchung über die Industriesektorenzugehörigkeit von Unternehmen in einer bestimmten Region lieferte das Ergebnis, dass 25 % der Unternehmen zum primären (P), 25 % zum sekundären (S) und 50 % zum tertiären Sektor (T) zählen. Auch wenn bei einem derartig einfach überschaubaren Resultat eigentlich keine grafische Darstellung erforderlich ist, wollen wir dieses einfache Beispiel im Folgenden dazu nutzen, die vorhergehend beschriebenen Darstellungsformen zu veranschaulichen:

#### 1. Piktogramm:

☺ ☺ ☺ ☺ ☺

Primärer Sektor

☺ ☺ ☺ ☺ ☺

Sekundärer Sektor

☺ ☺ ☺ ☺ ☺ ☺ ☺ ☺ ☺ ☺

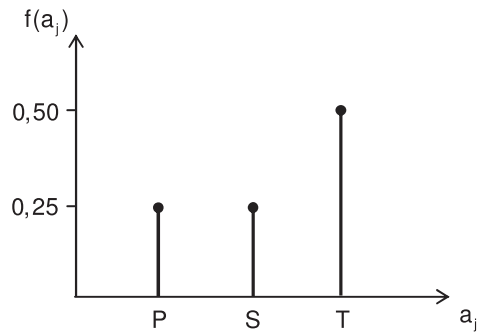
Tertiärer Sektor

Zur Konstruktion von Stab- und Säulendiagramm sowie des Häufigkeitspolygons im Koordinatensystem benötigen wir die Punkte

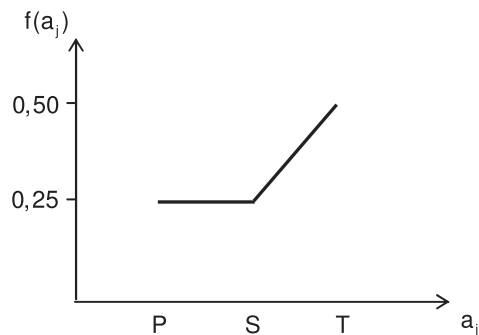
$$P_1(P; 0,25) \quad P_2(S; 0,25) \quad P_3(T; 0,50).$$

Wir erkennen daran, dass in derartigen grafischen Darstellungen nicht zwangsläufig Zahlenwerte auf der x-Achse abgetragen werden müssen.

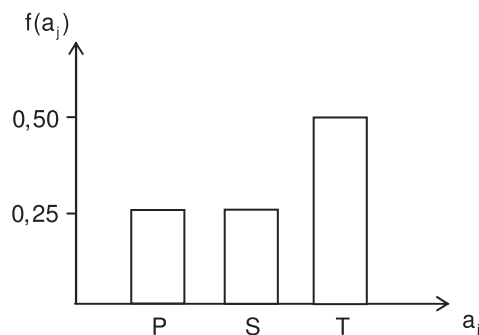
2. Stabdiagramm:



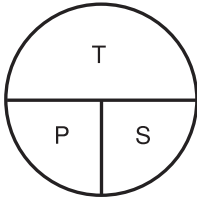
3. Häufigkeitspolygon:



4. Säulendiagramm:



## 5. Kreisdiagramm:



Winkelberechnung:

$$\alpha_1 = 360^\circ \cdot f(a_1) = 360^\circ \cdot 0,25 = 90^\circ$$

$$\alpha_2 = 360^\circ \cdot f(a_2) = 360^\circ \cdot 0,25 = 90^\circ$$

$$\alpha_3 = 360^\circ \cdot f(a_3) = 360^\circ \cdot 0,50 = 180^\circ$$

### 2.1.2 Empirische Verteilungsfunktion bei diskreten Merkmalen

Ist ein Merkmal  $X$  *ordinal- oder metrisch skaliert*, dann unterliegen seine Merkmalsausprägungen  $a_j$  einer *Rangordnung*. Es ist hier (nur hier) sinnvoll, die absoluten und relativen Häufigkeiten in aufsteigender Reihenfolge der  $a_j$  zu summieren.

Durch fortlaufende Aufsummierung (Kumulierung) der absoluten Häufigkeiten einer Häufigkeitsverteilung mit  $a_1 < a_2 < \dots < a_k$  lassen sich sog. *absolute Summenhäufigkeiten* oder *kumulierte absolute Häufigkeiten*

$$H(x) = H(X \leq x) = \sum_{a_j \leq x} h(a_j) \quad (12: 7)$$

ermitteln, wobei  $H(x)$  die Anzahl jener Merkmalsträger angibt, welche einen Merkmalswert besitzen, der höchstens (kleiner oder gleich)  $x$  beträgt. Falls  $x$  genau einem der  $a_j$  entspricht, sind also lediglich die absoluten Häufigkeiten bis zu dieser Ausprägung  $a_j$  aufzusummieren, d.h.  $H(a_j) = h(a_1) + h(a_2) + \dots + h(a_j)$ . Die ungewöhnliche Summenschreibweise in (12: 7) wird verwendet, da  $x$  auch Werte annehmen kann, für die kein  $a_j$  existiert. In einem solchen Fall werden die absoluten Häufigkeiten aufsummiert, die zu Merkmalswerten  $a_j$  gehören, die kleiner als  $x$  sind.

Analog lassen sich sog. *relative Summenhäufigkeiten* oder *kumulierte relative Häufigkeiten* ermitteln. Für diese gilt

$$F(x) = F(X \leq x) = \sum_{a_j \leq x} f(a_j), \quad (12: 8)$$

wobei  $F(x)$  den Anteil der Merkmalsträger angibt, welche einen Merkmalswert aufweisen, der höchstens  $x$  beträgt. Alternativ kann die Bestimmung von  $F(x)$  auch über den Zusammenhang

$$F(x) = \frac{H(x)}{n} \quad (12: 9)$$

erfolgen. Wir bezeichnen (12: 8) auch als **empirische Verteilungsfunktion**. Da für die Häufigkeitsverteilung  $a_1 < a_2 < \dots < a_k$  gilt, ist der Funktionswert für ein jedes  $x$ , welches kleiner als  $a_1$  ausfällt, unweigerlich Null. Es gibt schließlich keine Merkmalsträger, die eine Merkmalsausprägung aufweisen, die kleiner als  $a_1$  ist, da dies bereits der kleinste Wert im Datensatz ist. Der Funktionswert für ein jedes beliebiges  $x$ , das über dem größten Merkmalswert  $a_k$  liegt, ist gleich 1. Der Anteil der-

jeniger Merkmalsträger, die eine Merkmalsausprägung von weniger als  $x$  aufweisen, ist schließlich 100 %. Wir können also zusammenfassend festhalten, dass die empirische Verteilungsfunktion beschränkt ist, d.h.  $0 \leq F(x) \leq 1$  gilt. Wegen  $f(a_j) \geq 0$  gilt außerdem  $F(a_1) \leq \dots \leq F(a_j) \leq \dots \leq F(a_k)$ , d.h. die empirische Verteilungsfunktion ist monoton steigend. Wie wir im Beispiel I 2-3 sehen werden, hat sie außerdem grafisch die Gestalt einer **Treppenfunktion**, welche bei jedem  $a_j$  eine Sprungstelle mit einem Sprung in der Höhe von  $f(a_j)$  aufweist.

Bei vorliegenden kumulierten absoluten oder relativen Häufigkeiten können wir natürlich unmittelbar wieder auf die Einzelhäufigkeiten schließen. Es gelten nämlich die Zusammenhänge

$$h(a_j) = H(a_j) - H(a_{j-1}) \quad (I2: 10)$$

und

$$f(a_j) = F(a_j) - F(a_{j-1}). \quad (I2: 11)$$

### Beispiel I 2-3

### Empirische Verteilungsfunktion



In 100 Filialen einer Buchhandlung wurden an einem bestimmten Tag entweder 0, 1, 2 oder 3 Exemplare (Ausprägungen des Merkmals Absatzzahl) eines bestimmten Buches verkauft. Die resultierende Häufigkeitstabelle, die uns die Häufigkeit (absolut und relativ) der Filialen angibt, in denen es zu den jeweiligen Verkaufszahlen kam, zeigt sich wie folgt:

j	$a_j$	$h(a_j)$	$f(a_j)$
1	0	10	0,10
2	1	40	0,40
3	2	20	0,20
4	3	30	0,30
-	-	100	1,00

Ermitteln wir nun zunächst die absoluten und relativen kumulierten Häufigkeiten:

j	$a_j$	$h(a_j)$	$H(a_j)$	← Berechnung
1	0	10	10	$h(0)$
2	1	40	50	$h(0) + h(1)$
3	2	20	70	$h(0) + h(1) + h(2)$
4	3	30	100	$h(0) + h(1) + h(2) + h(3)$
-	-	100	-	-

j	$a_j$	$f(a_j)$	$F(a_j)$	← Berechnung
1	0	0,10	0,10	$f(0)$
2	1	0,40	0,50	$f(0) + f(1)$
3	2	0,20	0,70	$f(0) + f(1) + f(2)$
4	3	0,30	1,00	$f(0) + f(1) + f(2) + f(3)$
-	-	1,00	-	-

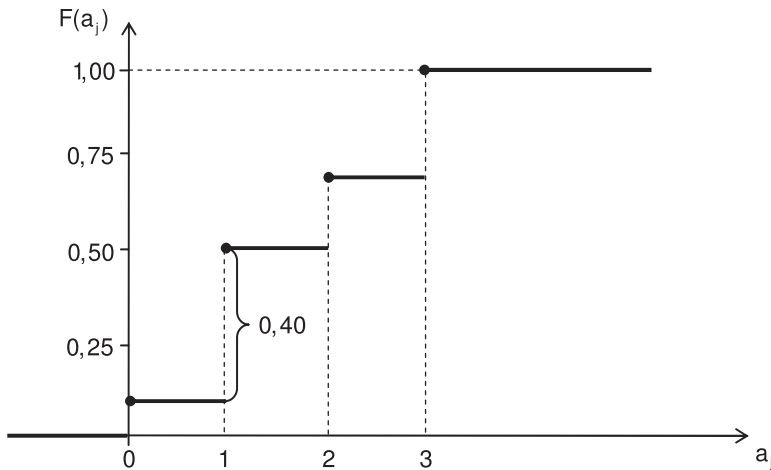
Zur besseren Veranschaulichung der Berechnung einzelner kumulierter Häufigkeiten wollen wir diese nun noch einmal für den Wert  $F(2)$  im Detail darlegen:

$$F(2) = f(0) + f(1) + f(2) = 0,1 + 0,4 + 0,2 = 0,7 \quad \text{oder alternativ} \quad F(2) = \frac{H(2)}{n} = \frac{70}{100} = 0,7$$

Soll nun beispielsweise der Wert von  $f(1)$  aus den relativen kumulierten Häufigkeiten errechnet werden, kann dies wie folgt geschehen:

$$f(1) = F(1) - F(0) = 0,50 - 0,10 = 0,40$$

Wie wir gleich sehen werden, lässt sich dieser Wert auch aus der grafischen Darstellung der empirischen Verteilungsfunktion ablesen, da in dieser die Sprunghöhe zwischen einzelnen Treppenstufen an einer Stelle  $a_j$  jeweils den zugehörigen Wert  $f(a_j)$  angibt. Für die grafische Konstruktion der empirischen Verteilungsfunktion tragen wir lediglich die Punkte  $P(a_j, F(a_j))$  in ein Koordinatensystem ein und versehen diese mit Linienzügen:



Auch die konkrete Funktionsschreibweise der empirischen Verteilungsfunktion können wir angeben. Außerdem können wir das in absoluten kumulierten Häufigkeiten formulierte Gegenstück zur empirischen Verteilungsfunktion bestimmen:

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0,10 & \text{" } 0 \leq x < 1 \\ 0,50 & \text{" } 1 \leq x < 2 \\ 0,70 & \text{" } 2 \leq x < 3 \\ 1,00 & \text{" } x \geq 3 \end{cases} \quad H(x) = \begin{cases} 0 & \text{für } x < 0 \\ 10 & \text{" } 0 \leq x < 1 \\ 50 & \text{" } 1 \leq x < 2 \\ 70 & \text{" } 2 \leq x < 3 \\ 100 & \text{" } x \geq 3 \end{cases}$$

Beantworten wir nun mittels dieser beiden Funktionen folgende Fragen:

- a) In wie viel Prozent der Filialen wurden höchstens 2 Bücher verkauft?

Die Antwort auf diese Fragestellung kann direkt als  $F(2) = 0,70$  bzw. als 70 % angegeben werden.

- b) In wie vielen Filialen wurde mindestens 1 Buch verkauft?

Hier können wir die Beziehung  $n - H(0) = 100 - 10 = 90$  nutzen, d.h. wir subtrahieren von der Gesamtzahl der Filialen die Anzahl derjenigen Filialen, in denen kein Buch verkauft wurde.



### 2.1.3 Klassierte Häufigkeitsverteilung bei stetigen Merkmalen

Liegt ein *stetiges* (oder *quasistetiges*) *Merkmal*  $X$  vor, so taucht das Problem auf, dass die Elemente der Urliste häufig alle voneinander verschieden sind. Das Aufstellen einer "regulären" Häufigkeitstabelle des im Abschnitt I 2.1.1 vorgestellten Typs bietet hier keinen Informationsgewinn, da die meisten Merkmalsausprägungen nur einmal vorkommen. Es empfiehlt sich in einem solchen Fall, die Ausprägungen von  $X$  in der Stichprobe in  $k$  **Klassen**  $K_j$  ( $j = 1, 2, \dots, k$ ) einzuteilen und die zugehörigen Klassenhäufigkeiten zu ermitteln. Die **absolute Klassenhäufigkeit**  $h_j$  ist dabei die Anzahl derjenigen Elemente, deren Merkmalswerte in die Klasse  $K_j$  fallen. Wir können also sagen, dass hier die Klassen  $K_j$  die Rolle der  $a_j$  aus der "regulären" Häufigkeitstabelle übernehmen. Die **relativen Klassenhäufigkeiten**  $f_j$  ergeben sich wieder über

$$f_j = \frac{h_j}{n} \quad (\text{I2: 12})$$

und weisen damit die gleichen Eigenschaften wie in (I2: 4) und (I2: 5) auf. Auch für die  $h_j$  gelten (I2: 1) und (I2: 2) analog. Zudem sei erwähnt, dass wir auch *kumulierte Klassenhäufigkeiten* gemäß der unter I 2.1.2 beschriebenen Vorgehensweise bestimmen können.

Wieviele Klassen sinnvollerweise gebildet werden sollten, kann aus DIN 55302 (Blatt 1) entnommen werden. Hier finden sich unter anderem die Empfehlungen bei Stichprobenumfängen  $n$  von 100, 1.000 und 10.000 eine Anzahl von 10, 13 bzw. 16 Klassen zu wählen. Natürlich kann aus Gründen der Übersichtlichkeit der Darstellung auch bereits bei geringerem  $n$  eine Klassierung vorgenommen werden. Außerdem kann aufgrund des Untersuchungsobjekts auch eine andere Klassenanzahl geeigneter erscheinen.

Um Probleme bei der Zuordnung einzelner Merkmalswerte zu den Klassen zu vermeiden, empfiehlt es sich die Klassen  $K_j$  in der Form  $a_j^u < x \leq a_j^o$  bzw.  $]a_j^u, a_j^o]$  zu definieren. Jede Klasse  $K_j$  ist damit gekennzeichnet durch eine **Klassenuntergrenze**  $a_j^u$  und eine **Klassenobergrenze**  $a_j^o$ . Da die einzelnen Klassen direkt ineinander übergehen, entspricht bei  $k$  Klassen die Untergrenze einer Klasse jeweils der Obergrenze der vorhergehenden Klasse.

$$a_j^u = a_{j-1}^o \quad (j = 2, 3, \dots, k) \quad (\text{I2: 13})$$

Die Untergrenze der ersten Klasse können wir anhand des kleinsten Merkmalswerts der Stichprobe wählen. Häufig wird sie auch auf Null gesetzt. Die letzte Klasse  $K_k$  besitzt häufig eine Form  $a_k^u < x < \infty$  bzw.  $]a_k^u, \infty[$  und wird als *offene Randklasse* bezeichnet. Dies bedeutet, dass in diese Klasse alle Merkmalsausprägungen fallen, die größer als  $a_k^u$  sind.

Die sog. **Klassenbreite**  $\Delta a_j$  der  $j$ -ten Klasse, die nach Möglichkeit durchgehend einheitlich (*konstant*) gestaltet werden sollte, ist (außer für offene Randklassen) als

$$\Delta a_j = a_j^o - a_j^u \quad (j = 1, 2, \dots, k) \quad (\text{I2: 14})$$

definiert. Bei der Wahl konstanter Klassenbreiten sind jedoch häufig manche Klassen sehr schwach besetzt. Dies trifft vor allem auf die Randklassen zu. Dann ist es evtl. sinnvoll, diese Klassen breiter zu wählen.

Besonders für die im Kapitel I 2.2 behandelnden Messzahlen ist es von entscheidender Wichtigkeit, jeder Klasse  $K_j$  einen repräsentativen Zahlenwert zuzuordnen. Dazu bietet sich die **Klassenmitte**  $a_j^m$  an, für welche (außer bei offenen Randklassen) folgendes gilt:

$$a_j^m = \frac{1}{2} \cdot (a_j^u + a_j^o) \quad (j = 1, 2, \dots, k) \tag{I2: 15}$$

Die Kombination der Klassen  $K_j$  mit ihren Klassenhäufigkeiten  $h_j$  bzw.  $f_j$  bildet die **klassierte Häufigkeitsverteilung** des Merkmals X. Wieder haben wir die Möglichkeit einer tabellarischen Darstellung. Abbildung I 6 veranschaulicht eine typische *klassierte Häufigkeitstabelle*.

Klassennummer	Klasse $K_j$	Klassenbreite	Klassenmitte	absolute Klassenhäufigkeit	relative Klassenhäufigkeit
j	über $a_j^u$ bis $a_j^o$	$\Delta a_j$	$a_j^m$	$h_j$	$f_j$
1	$[a_1^u, a_1^o]$	$\Delta a_1$	$a_1^m$	$h_1$	$f_1$
2	$[a_1^o, a_2^o]$	$\Delta a_2$	$a_2^m$	$h_2$	$f_2$
...	...	...	...	...	...
k	$[a_{k-1}^o, a_k^o]$	$\Delta a_k$	$a_k^m$	$h_k$	$f_k$
$\Sigma$	-	-	-	n	1

Abbildung I 6: Häufigkeitstabelle bei klassierten Daten



**Beispiel I 2-4**

Häufigkeitstabelle bei klassierten Daten

Nehmen wir an, eine Erhebung in 47 amerikanischen Staaten ergab folgende Anzahlen polizeilich gemeldeter Straftaten bezogen auf 100.000 Einwohner:

79,3 163,4 57,8 196,1 123,5 68,1 96,3 155,8 85,3 70,9 167,4 84,8 51,2 66,4 79,1  
 94,3 53,1 92,4 75,1 122,5 74,8 43,3 121,9 96,5 52,7 199,1 34,2 121,6 104,5 69,3  
 37,7 75,9 107,2 92,4 65,1 127,6 83,3 56,5 82,8 115,8 88,1 54,2 82,4 103,0 44,5  
 50,8 84,9

Eine erste Untersuchung des Datenmaterials zeigt, dass die kleinste Kriminalitätsrate bei 34,2 und die größte bei 199,1 liegt. Die Mehrzahl der Staaten verzeichnete Kriminalitätsraten im Bereich von 50 bis 120. Um sich einen genaueren Überblick zu verschaffen, bietet sich hier die Aufstellung einer klassierten Häufigkeitstabelle mit Klassenbreiten von 25 an. Da dabei die entstehende untere Randklasse jedoch nicht besetzt wäre, verdoppeln wir diese und erhalten damit folgende Tabelle (Achtung: Rundungsungenauigkeiten bei relativen Häufigkeiten!):



<http://www.springer.com/978-3-658-06438-9>

Statistik und Ökonometrie für Wirtschaftswissenschaftler

Eine anwendungsorientierte Einführung

Auer, B.R.; Rottmann, H.

2015, XXV, 763 S. 108 Abb., Softcover

ISBN: 978-3-658-06438-9