

2 Statistische Maßzahlen

Übersicht

2.1	Quantile, speziell Median, Quartile und Perzentile	25
2.2	Modus, Median, arithmetisches Mittel	28
2.3	Arithmetisches, geometrisches, harmonisches Mittel	31
2.4	Bedeutung und Einteilung der Streuungsmaße	37
2.5	Absolute Streuungsmaße	41
2.6	Dimensionslose Streuungsmaße: Nominalabweichung und relative Streuungsmaße	45

2.1 Quantile, speziell Median, Quartile und Perzentile

Für die Bestimmung von Quantilen ist mindestens das Skalenniveau einer Ordinalskala notwendig. Bei einer Ordinalskala können die Merkmalsausprägungen in eine natürliche Reihenfolge gebracht werden, wie dies zum Beispiel bei der Güteklasse von Obst möglich ist.

Die **Urliste** ist das direkte Ergebnis einer Datenerhebung, also die ursprüngliche Aufzeichnung der Beobachtungs- oder Messwerte. Die Urliste wird mit

$$x_1, x_2, x_3, \dots, x_n$$

bezeichnet, wobei n für die Anzahl der Beobachtungen (Stichprobenumfang) steht.

Werden die Merkmalsausprägungen der Größe nach sortiert, so erhält man die **geordnete Stichprobe (Rangliste)**. $x_{[1]}$ bezeichnet den kleinsten Wert, $x_{[2]}$ den zweitkleinsten, bis schließlich $x_{[n]}$ den größten Wert definiert. $x_{[k]}$ heißt die k te Ordnungsgröße.

Anhand einer Rangliste können für vorliegende Merkmalsausprägungen Quantile bestimmt werden.

Ein **p -Quantil** (Q_p ; manchmal auch noch als „Fraktil“ bezeichnet) ist jener Beobachtungswert, der größer oder gleich als mindestens $100 \cdot p$ % der Werte und zugleich kleiner oder gleich als $100 \cdot (1 - p)$ % der Werte ist.

Beispiel 2.1

Das 0,5-Quantil ist somit der Beobachtungswert, der größer oder gleich 50 % der beobachteten Werte und zugleich kleiner oder gleich als 50 % der Werte ist. Das 0,5-Quantil wird auch Median ($\tilde{x}$) oder 2. Quartil genannt, das 0,25-Quantil wird 1. Quartil oder unteres Quartil (Q_1) und das 0,75-Quantil 3. Quartil oder oberes Quartil (Q_3) genannt.

Das untere (obere) Quartil teilt die Rangliste im Verhältnis 1 : 3 (3 : 1). Der Median wird auch häufig Zentralwert genannt oder der Wert „in der Mitte“. Er definiert die „Mitte“ der Rangreihe (geordnete Stichprobe).

Wie der Name andeutet, bilden Quartile Schnittpunkte zwischen den Vierteln einer Verteilung (lateinisch *quartarius* = Viertel, *quartus* = der Vierte). ■

Die gebräuchlichste Berechnungsvariante von Quantilen lässt sich folgendermaßen beschreiben: Wird das p -Quantil aus einer Reihe von n Messwerten gesucht, so muss zunächst die Rangliste dieser n Messwerte erstellt werden. Danach wird das Produkt aus n und p gebildet und das p -Quantil nach einer der folgenden Regeln bestimmt.

Ist $n \cdot p$ keine ganze Zahl: Runde $n \cdot p$ auf die nächste größere Zahl. Die daraus resultierende Zahl gibt direkt die Stelle an, wo sich das p -Quantil in der Rangreihe finden lässt.	Ist $n \cdot p$ eine ganze Zahl: Diese Zahl gibt die Stelle an, die zunächst in der Rangreihe gesucht werden muss. Der Mittelwert aus dem so gefundenen Wert und dem darauf folgenden Wert ist das gesuchte p -Quantil.
--	---

Quantile gehören zu den Lagemaßen (Lokalisationsmaße). Mit Hilfe von Lagemaßen kann angegeben werden, in welchen Bereichen der Messskala die Beobachtungseinheiten zu liegen kommen. Die Lagemaße geben die zentrale Tendenz der Daten wieder.

Beispiel 2.2

Aus den Körpergrößen von 24 Frauen ergab sich die folgende Rangliste:

153 154 158 159 159 160 161 161 162 163 165 166
167 168 170 174 174 175 176 178 178 179 181 182

Um den Median zu berechnen, wird das Produkt aus $n = 24$ und $p = 0,5$ gebildet; das Ergebnis lautet 12 ($= 24 \cdot 0,5$). 12 ist eine ganze Zahl, daher ergibt sich der Median $\tilde{x}$ aus dem Mittel aus $x_{[12]} = 166$ und $x_{[13]} = 167$, also zu 166,5 ($= (166 + 167)/2$). Das untere Quartil Q_1 errechnet sich aus dem Produkt von $n = 24$ und $p = 0,25$ und liefert 6, eine ganze Zahl. Daher bestimmt sich das untere Quartil Q_1 zu 160,5 ($= (x_{[6]} + x_{[7]})/2 = (160 + 161)/2$) und das obere Quartil Q_3 ergibt sich aus dem Mittel von $x_{[18]} = 175$ und $x_{[19]} = 176$, also 175,5. ■

Wie Beispiel 2.2 zeigt, gibt es bei der Berechnung mit einem geraden Stichprobenumfang n kein einzelnes mittleres Element, sondern zwei. Laut der Definition des Medians

wären sowohl die Werte der beiden mittleren Merkmalsausprägungen als auch alle Werte dazwischen ein Median, als Konvention hat sich aber im Allgemeinen der Mittelwert aus den beiden Werten herauskristallisiert. Ist jedoch der Stichprobenumfang n ungerade, so liegt der Median genau in der Mitte der Rangliste.

Der Median kann auch direkt bei ordinalskalierten Merkmalen, die anhand einer Häufigkeitstabelle vorliegen, berechnet werden. In einem Fragebogen zur Krankheitsverarbeitung sollten 160 Patienten auf einer Fünferskala angeben, inwieweit sie aktive Anstrengungen zur Lösung ihrer gesundheitlichen Probleme unternehmen. Die entsprechenden Häufigkeiten sind in Tab. 2.1 wiedergegeben.

Tab. 2.1: Absolute und kumulierte Häufigkeiten bei 160 Patienten zur Lösung gesundheitlicher Probleme

Anstrengung zur Lösung gesundheitlicher Probleme	absolute Häufigkeit	kumulierte absolute Häufigkeit
gar nicht	12	12
wenig	25	37
mittelmäßig	23	60
ziemlich	53	113
sehr stark	47	160

Der Begriff **absolute Häufigkeit** ist gleichbedeutend mit dem „umgangssprachlichen“ Begriff Anzahl. Die absolute Häufigkeit ist die Zählung, wie oft zum Beispiel von den 160 Patienten diese Frage nach ihrer Anstrengung zur Lösung gesundheitlicher Probleme mit „wenig“ beantwortet wurde. Laut der Tab. 2.1 waren dies 25 Personen. Die **kumulierte Häufigkeit** (auch Summenhäufigkeit) berechnet sich als Summe der absoluten Häufigkeiten der Merkmalsausprägungen von der kleinsten bis zur jeweiligen Merkmalsausprägung. So ergibt sich die kumulierte absolute Häufigkeit für „mittelmäßig“ aus der Summe von 12, 25 und 23, also zu 60. Es ist darauf zu achten, dass bei der Häufigkeitstabelle die Merkmalsausprägungen der Rangliste entsprechend geordnet sind.

Nach den erläuterten Regeln zur Bestimmung des Medians ergibt sich bei der vorliegenden Frage an die 160 Patienten „ziemlich“. Bei insgesamt 160 Werten liegt der Median nämlich zwischen dem 80. und 81. Ranglistenelement. Die kumulierte absolute Häufigkeit zeigt an, dass sowohl der 80. als auch der 81. Wert „ziemlich“ ist, womit auch der Median diese Merkmalsausprägung annimmt.

Quartile teilen die Merkmalsausprägungen in vier gleich große Teile (in Bezug auf den Umfang der Merkmalsausprägungen); die Abb. 2.1 veranschaulicht diese Tatsache.

Die **Dezile** (lateinisch, dt. „Zehntelwerte“) zerlegen die Rangliste in zehn gleich große Teile (bezüglich Umfang). Das 1. Dezil (unteres Dezil) gibt an, welcher Wert die unteren 10 % von den oberen 90 % der Rangliste trennt, das 2. Dezil, welcher Wert die unteren

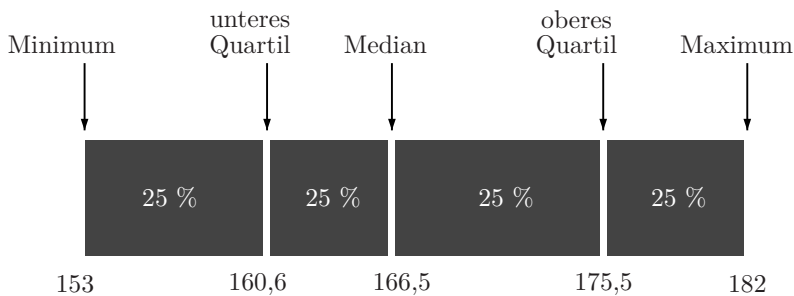


Abb. 2.1: Vierteilung der Körpergröße von Frauen ($n = 24$)

20 % von den oberen 80 % der Rangliste trennt, ... und das 9. Dezil (oberes Dezil) gibt an, welcher Wert die unteren 90 % von den oberen 10 % der Rangliste trennt.

Beispiel 2.3

Anhand des Beispiels 2.2 werden nun die Dezile berechnet. Das erste (oder unteres) Dezil (D_1) ergibt sich aus $24 \cdot 0,1 = 2,4$, also die Stelle 3, das untere Dezil ist daher $x_{[3]} = 158$. Das neunte (oder oberes) Dezil (D_9) ergibt sich aus $24 \cdot 0,9 = 21,6$, also die Stelle 22, das obere Dezil ist daher $x_{[22]} = 179$. Alle Dezile sind gegeben durch 158; 159; 161; 163; 166,5; 170; 174; 178; 179. ■

Durch **Perzentile** (lateinisch, dt. „Hundertstelwerte“), auch Prozentränge genannt, wird die Rangliste in hundert gleich große Teile zerlegt. Perzentile teilen die Rangliste also in 1 %-Segmente auf. Daher können Perzentile als p -Quantile betrachtet werden, bei denen $100 \cdot p$ eine ganze Zahl ist. So entspricht das $Q_{0,95}$ dem Perzentil P_{95} : Unterhalb dieses Punktes liegen 95 % aller Merkmalsausprägungen der Rangliste.

2.2 Modus, Median, arithmetisches Mittel

Der Modalwert ist besonders zur Beschreibung der Lage nominalskaliertter Merkmale, wie zum Beispiel dem Familienstand, geeignet. Er gibt die zentrale Tendenz an.

Als **Modalwert** (Modus) x_M einer Menge von Messwerten wird der am häufigsten auftretende Wert bezeichnet. Wenn zwei oder mehr Werte am häufigsten vorkommen, gibt es mehrere Modalwerte.

Der Modalwert kann ebenfalls zur Beschreibung der Lage ordinalskaliertter (nur die Ordnung der Merkmalsausprägungen spielt eine Rolle) beziehungsweise metrischskaliertter Merkmale (jene Merkmale, die durch Messen, Zählen oder Wiegen bestimmt werden können) eingesetzt werden.

Allerdings ist bei unklassifizierten metrischen Merkmalen die Angabe des Modalwerts oft wenig sinnvoll. Wenn die einzelnen Merkmalsausprägungen nur in geringer Anzahl

vorliegen, wird als Modalwert x_M der mittlere Wert der am häufigsten auftretenden Klasse beziehungsweise den am häufigsten auftretenden Klassen verwendet (siehe dazu Beispiel 2.7). Unter einer Klassifizierung (von lateinischen *classis*, dt. Klasse) ist das Zusammenfassen von Merkmalsausprägungen zu Klassen zu verstehen.

Beispiel 2.4

Tabelle 2.2 zeigt die Häufigkeitstabelle einer Studie, bei welcher der Familienstand abgefragt wurde (Stichprobenumfang $n = 97$).

Tab. 2.2: Häufigkeitstabelle für das Merkmal „Familienstand“

Familienstand	absolute Häufigkeit	relative Häufigkeit	prozentuelle Häufigkeit
ledig	28	0,289	28,9 %
verheiratet	43	0,443	44,3 %
verwitwet	11	0,113	11,3 %
geschieden	15	0,155	15,5 %
gesamt	97	1	100 %

Die „absolute Häufigkeit“ ergibt sich aus der Anzahl, wie oft die entsprechende Merkmalsausprägung des Merkmals Familienstand vorkommt. So kam zum Beispiel bei den 97 Befragten 11-mal verwitwet und 15-mal geschieden vor. Die „**relative Häufigkeit**“ (der Anteil) ergibt sich mittels Division der absoluten Häufigkeiten durch den Stichprobenumfang n . So ist zum Beispiel die relative Häufigkeit von ledig $28/97 = 0,289$. Die **prozentuelle Häufigkeit** ergibt sich aus der relativen Häufigkeit multipliziert mit 100. Die Unterschiedlichkeit der einzelnen Häufigkeiten wird anschließend nochmals verdeutlicht.

Es wurden 97 Personen nach dem Familienstand gefragt. Bei der Auszählung wurde festgestellt, dass 28 Personen in die Klasse „ledig“ fallen.

- Die absolute Häufigkeit von ledig ist daher 28.
- Die relative Häufigkeit dieser Klasse ist 28 zu 97 ($28/97 = 0,289$).

Das heißt, 28,9 % (die prozentuelle Häufigkeit) der Befragten sind ledig. Die relative Häufigkeit ist die absolute Häufigkeit von ledig „relativ“ zur Anzahl der befragten Personen. ■

Die n Beobachtungswerte $x_1, x_2, x_3, \dots, x_n$ werden Urliste eines metrisch skalierten Merkmals genannt, also Merkmale, die zähl-, wieg- beziehungsweise messbar sind. n wird auch der Stichprobenumfang (Anzahl der Beobachtungen) genannt.

Das **arithmetische Mittel** $\bar{x}$, ein Lagemaß, welches die zentrale Tendenz angibt, ist gegeben durch:

$$\bar{x} = \frac{1}{n} \cdot (x_1 + x_2 + x_3 + \dots + x_n)$$

Die dem arithmetischen Mittel zugrunde liegende Idee ist, dass als Stellvertreter aller aufgetretenen Daten jene Zahl gewählt wird, die sich bei einer gleichmäßigen Aufteilung der Summe aller Daten (genannt die Merkmalssumme) auf die Beobachtungseinheiten ergeben würde. Das arithmetische Mittel der Einkommen ist also jenes, das auf jeden Einzelnen fallen würde, wenn das gesamte Einkommen aller Personen gleichmäßig auf alle Personen aufgeteilt werden würde.

Beispiel 2.5

Im Beispiel der Körpergröße von Frauen ($n = 24$) lautete die Urliste:

167	170	178	154	176	162	182	166	153	165	161	175
159	168	159	158	179	174	181	174	163	160	161	178

Das arithmetische Mittel ergibt sich demnach zu

$$\bar{x} = \frac{1}{24} \cdot (167 + 170 + 154 + 176 + 162 + \cdots + 161 + 178) = 167,6$$

■

Das arithmetische Mittel ist empfindlich gegenüber extremen Werten (Ausreißern) und daher für mehrgipfelige und sehr schiefe Histogramme nicht geeignet (vergleiche Abschn. 4.3).

Im Falle diskreter Merkmale hat das arithmetische Mittel möglicherweise keine Entsprechung in der Wirklichkeit (zum Beispiel durchschnittliche Kinderzahl einer Frau von 1,75).

Der Median (der Wert in der „Mitte“) ist unempfindlich gegenüber Ausreißern (der Median ist sozusagen robust), jedoch für mehrgipfelige Histogramme ungeeignet.

Beispiel 2.6

Als Beispiel für die Empfindlichkeit des arithmetischen Mittels in Bezug auf Ausreißer im Vergleich zum Median wird das Beispiel 2.5 der Körpergrößen von Frauen ($n = 24$) betrachtet:

Rangliste:	153	154	158	159	159	160	161	161	162	163	165	166
	167	168	170	174	174	175	176	178	178	179	181	182

Der Median $\tilde{x}$, der Wert in der Mitte, ergibt sich aus dem Mittel des 12. und 13. Ranglistenelements, in diesem Beispiel aus $x_{[12]} = 166$ und $x_{[13]} = 167$, also zu 166,5; das arithmetische Mittel war 167,6.

Ist nun bei der Übertragung der Daten ein Fehler entstanden und wurde etwa 253 statt 153 notiert, so ändert sich der Median $\tilde{x}$ zu 167,5, das arithmetische Mittel $\bar{x}$ ändert sich jedoch auf 171,8. ■

Beispiel 2.7

Soll aus dem Beispiel 2.6 der Modalwert x_M für das unklassifizierte metrische Merkmal „Körpergröße“ berechnet werden, so wäre jener Wert, der am häufigsten vorkommt, nicht eindeutig. Die Werte 159, 161, 174 und 178 kommen jeweils zweimal vor, sind also alle der Modalwert x_M . Als Grundlage für die Klasseneinteilung dient meist die Rangliste der Merkmalsausprägungen.

Um eine Klassifikation vorzunehmen, müssen zunächst sinnvolle Klassengrenzen beziehungsweise Wertebereiche für die einzelnen Klassen festgelegt werden. Dies geschieht zumeist durch die Angabe von Intervallen. Hierbei wird eine mathematische Kurzschreibweise angewendet: Statt 4,01 bis 6,00 wird lediglich $(4; 6]$ geschrieben. Durch diese spezielle Verwendung von runden und eckigen Klammern ist klar festgelegt, in welche Klasse ein Wert fällt, der genau der Klassengrenze entspricht. Eine runde Klammer gibt an, dass dieser Wert gerade nicht mehr zu dieser Klasse gehört. Eine eckige Klammer schließt den Wert in die Klasse ein. Das Ergebnis der Klasseneinteilung kann mit Hilfe einer Tabelle, welche die Klassenhäufigkeiten angibt, dargestellt werden.

Als Faustregel für die Anzahl der Klassen hat sich die Berechnungsmethode der Quadratwurzel des Stichprobenumfangs n durchgesetzt, also $\approx \sqrt{n}$.

In unserem Beispiel ist $n = 24$, damit ergibt sich eine Klassenanzahl von 5. Die Breite der Klasse kann mit Hilfe der Differenz des größten und des kleinsten Wertes, also $x_{[n]} - x_{[1]}$ berechnet werden. Diese Größe wird auch Spannweite (Range) R genannt und ist ein so genannter Streuungsparameter, vergleiche auch Abschn. 2.4 und Abschn. 2.5. In unserem Beispiel ergibt sich die Spannweite $R = 182 - 153 = 29$. Die Klassenbreite ergibt sich nun aus der Spannweite dividiert durch die Anzahl der Klassen, demnach zu $29/5$, also etwa 6.

Aufgrund dieser Informationen lässt sich eine sinnvolle Klassifizierung der Körpergröße von Frauen ($n = 24$) bestimmen, nämlich die folgenden Klassen.

$(152; 158]$ $(158; 164]$ $(164; 170]$ $(170; 176]$ $(176; 182]$

Anhand dieser Klasseneinteilung des unklassifizierten metrischen Merkmals der Körpergröße von Frauen ergibt sich, dass die häufigsten Werte in der Klasse $(158; 164]$ liegen; Die Klassenhäufigkeit beträgt 7. Der Modalwert x_M ist die Klasse $(158; 164]$ beziehungsweise der mittlere Wert dieser Klasse, also $x_M = 161$. ■

2.3 Arithmetisches, geometrisches, harmonisches Mittel

Trotz seines breiten Anwendungsbereichs gibt das arithmetische Mittel $\bar{x}$, umgangssprachlich oft „Durchschnitt“ genannt, bei bestimmten Merkmalen aus sachlogischen Gründen nicht den richtigen Durchschnitt an. Dies ist dann der Fall, wenn relative

Änderungen als Merkmalsausprägung von Interesse sind, wie dies bei zeitabhängigen Messzahlen der Fall ist. Zeitabhängige Messzahlen sind zu erhalten, indem zwei Beobachtungen mit unterschiedlichem Zeitbezug, aber für dasselbe Merkmal ins Verhältnis gesetzt werden. Das **geometrische Mittel** $\bar{x}_g$ wird berechnet, wenn eine durchschnittliche relative Änderung berechnet werden soll. Die relative Änderung kann als Quotient von End- und Anfangswert eines Zeitraums dargestellt werden beziehungsweise es wird der Endwert über das Intervall durch das Produkt von Anfangswert und den bis zum Endzeitpunkt angefallenen n relativen Änderungen beschrieben. In solchen Fällen wird die n -te Wurzel aus dem Produkt all dieser n Werte $(x_1, x_2, x_3, \dots, x_n)$ gebildet, wobei vorausgesetzt wird, dass alle Werte positiv sind ($x_i > 0$ für $1 \leq i \leq n$) – also ist dies nur bei verhältnisorientierten Merkmalen sinnvoll.

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

Wenn die Definition des geometrischen Mittels hoch n genommen wird

$$\begin{aligned}\bar{x}_g^n &= x_1 \cdot x_2 \cdot x_3 \cdots x_n \\ 1 &= \frac{x_1}{\bar{x}_g} \cdot \frac{x_2}{\bar{x}_g} \cdot \frac{x_3}{\bar{x}_g} \cdots \frac{x_n}{\bar{x}_g}\end{aligned}$$

sehen wir, dass das geometrische Mittel in einer besonderen Weise „zentral“ ist: Die n Quotienten $\frac{x_i}{\bar{x}_g}$, von denen einige kleiner als Eins, andere größer als Eins sind, multiplizieren sich gerade zu Eins auf. Wegen dieser Eigenschaft ist das geometrische Mittel eher geeignet, Quotienten, Prozente und Wachstumsraten zu ermitteln, wenn das arithmetische Mittel versagen würde. Aus der Definition des geometrischen Mittels ergibt sich durch Logarithmieren:

$$\ln \bar{x}_g = \frac{1}{n} \sum_{i=1}^n \ln x_i$$

Das geometrische Mittel entspricht also dem arithmetischen Mittel der Logarithmen.

Das geometrische Mittel $\bar{x}_g$ findet zum Beispiel bei „durchschnittlichen“ Gehaltserhöhungen, Steigerung des Bruttosozialprodukts oder bei Durchschnittsverzinsungen Anwendung. Das geometrische Mittel ist immer kleiner oder gleich dem arithmetischen Mittel.

Zur Veranschaulichung der Anwendungsfelder des arithmetischen Mittels und des geometrischen Mittels dient nachfolgendes Beispiel.

Beispiel 2.8

Tabelle 2.3 zeigt die Umsatzsteigerung einer Boutique seit ihrer Eröffnung im Jahre 2005.

Die Fragen nach der durchschnittlichen Umsatzhöhe und der durchschnittlichen Umsatzsteigerung sollen beantwortet werden. Die durchschnittliche Umsatzhöhe wird nach dem arithmetischen Mittel $\bar{x}$ zu

$$\bar{x} = \frac{1}{5} \cdot (300.000 + 315.000 + 378.000 + 415.800 + 478.170) = 377.394$$

errechnet.

Tab. 2.3: Umsatzentwicklung einer Boutique für die Jahre 2005 bis 2009 in 1000 Euro

Jahr	Umsatz in 1000 Euro
2005	300,00
2006	315,00
2007	378,00
2008	415,80
2009	478,17

Die Frage nach der durchschnittlichen Umsatzsteigerung kann nur mit dem geometrischen Mittel beantwortet werden. Der Umsatz der Boutique lag im Jahr

2005 bei 300.000 Euro; er wurde im Jahr

2006 um 5 % auf $300000 \cdot 1,05 = 315.000$ Euro,

2007 um 20 % auf $300000 \cdot 1,05 \cdot 1,20 = 378.000$ Euro,

2008 um 10 % auf $300000 \cdot 1,05 \cdot 1,20 \cdot 1,10 = 415.800$ Euro,

2009 um 15 % auf $300000 \cdot 1,05 \cdot 1,20 \cdot 1,10 \cdot 1,15 = 478.170$ Euro gesteigert.

Die Tab. 2.4 zeigt die relativen Umsatzänderungen:

Tab. 2.4: Relative Umsatzänderung einer Boutique für die Jahre 2005 bis 2009

Jahr	Umsatz in 1000 Euro	relative Umsatzänderung
2005	300,00	-
2006	315,00	1,05
2007	378,00	1,20
2008	415,80	1,10
2009	478,17	1,15

Die Gesamtsteigerung des Umsatzes von 2005 bis 2009 wird bestimmt durch das Produkt $x_1 \cdot x_2 \cdot x_3 \cdot x_4 = 1,05 \cdot 1,20 \cdot 1,10 \cdot 1,15$. Aus diesem Grund ist es wenig sinnvoll, das arithmetische Mittel, das auf einer Summe basiert, zur Berechnung der durchschnittlichen Umsatzsteigerung heranzuziehen. Für die Boutique ergibt sich ($n = 4$):

$$\bar{x}_g = \sqrt[4]{1,05 \cdot 1,20 \cdot 1,10 \cdot 1,15} = 1,1236$$

Das heißt, die durchschnittliche relative Umsatzsteigerung von 1,1236 ist gleich einer durchschnittlichen Umsatzsteigerung von 12,36 %. Das Beispiel zeigt, dass nicht die Wachstumsraten (zum Beispiel 20 %) verwenden werden dürfen, sondern stattdessen die

entsprechenden Wachstumsfaktoren ($1,20 = 1 + 20\%$). Umgekehrt ist dann vom erhaltenden Ergebnis wieder 1 abzuziehen und der Wert 0,1236 mit seiner Äquivalenz 12,36 % anzugeben. Eine durchschnittliche Umsatzsteigerung von 12,36 % bedeutet, dass die Boutique den Umsatz 478170 im Jahr 2009 erreichen würde, wenn im Jahr 2005 der Umsatz 300000 ist und die Umsatzsteigerung pro Jahr genau 12,36 % beträgt.

In diesem Fall kann jedoch die durchschnittliche Umsatzsteigerung leichter durch die vierte Wurzel des Quotienten aus Endwert und Anfangswert berechnet werden. Sie ergibt sich zu

$$\bar{x}_g = \sqrt[4]{\frac{478170}{300000}} = 1,1236.$$

Achtung: Angenommen, es ergibt sich aufgrund einer Prognose im Jahr 2010 ein Umsatzrückgang von 15 % (dies ist ein Minus von 15 %), dann wäre die relative Änderung von 2009 auf 2010: 0,85. Bei einem Umsatzwachstum von 200 % im Jahr 2010 wäre die relative Änderung von 2004 auf 2005: 3. ■

Beispiel 2.9

Im Laufe von vier Jahren betrug der Kurs für eine Aktie jeweils zu Jahresbeginn 500, 600, 400 und 700 Euro. Eine Person tätigte zu diesen Kursen zu jedem Jahresbeginn Aktienkäufe. Berechnen Sie den durchschnittlichen Kaufkurs (pro Aktie), falls

1. jeweils 100 Stück
2. jeweils für den gleichen Betrag von 84.000 Euro

Aktien gekauft werden. Durch welche Mittelwertbildung kann der jeweilige durchschnittliche Kaufkurs berechnet werden?

Bei 1. ergibt sich die Mittelwertbildung folgendermaßen:

$$\begin{aligned}\bar{x} &= \frac{1}{400} \cdot (100 \cdot 500 + 100 \cdot 600 + 100 \cdot 400 + 100 \cdot 700) \\ &= \frac{1}{4} \cdot (500 + 600 + 400 + 700) \\ &= 550 \text{ Euro}\end{aligned}$$

550 Euro lautet das arithmetische Mittel der vier Einzelkurse.

Bei 2. ergibt sich der Gesamtbetrag zu $4 \cdot 84.000 = 336.000$. Die Anzahl der gekauften Aktien ist

$$\frac{84.000}{500} + \frac{84.000}{600} + \frac{84.000}{400} + \frac{84.000}{700} = 168 + 140 + 210 + 120 = 638.$$

Der durchschnittliche Kaufkurs ist daher

$$\frac{336.000}{638} = 526,65 \text{ Euro pro Aktie.}$$

Dieser Wert wäre auch durch folgende Berechnung zu erhalten gewesen:

$$526,65 = \frac{4}{\frac{1}{500} + \frac{1}{600} + \frac{1}{400} + \frac{1}{700}}$$

526,65 Euro ist das harmonische Mittel der vier Einzelkurse. ■

Das harmonische Mittel $\bar{x}_h$ ist ein geeignetes Lagemaß für Größen, die durch einen Bezug auf eine Einheit definiert sind (das heißt, die Zählvariable in den Einzelwerten ist konstant), zum Beispiel von Geschwindigkeiten (Strecke pro Zeiteinheit) oder Ernteträgen (Gewicht oder Volumen pro Flächeneinheit). Das harmonische Mittel $\bar{x}_h$ darf nur für Merkmale berechnet werden, bei denen mindestens die Quotienten zweier Merkmalsausprägungen sinnvoll interpretiert werden können, das Merkmal muss also mindestens auf einer Verhältnisskala vorliegen.

Eine komplizierte physikalische Dimension liegt vor, wenn diese beispielsweise der Quotient mehrerer physikalischer Maßeinheiten ist (zum Beispiel [1/h], [km/h], [Euro/kg]). Solche Quotienten zweier Merkmale werden Verhältniszahlen genannt. Einfache physikalische Dimensionen (zum Beispiel [1], [km], [kg], [Euro]) lassen sich als komplizierte Dimensionen auffassen, wenn diese durch die leere Dimension [1] dividiert werden (zum Beispiel [1/1], [km/1], [kg/1], [Euro/1]). Stimmen die Dimensionen der Nenner mit denen der absoluten Häufigkeiten überein, ist das arithmetische Mittel zu bilden. Liegt diese Gleichwertigkeit nicht vor, sondern stimmen die Dimensionen des Zählers mit denen der absoluten Häufigkeiten überein, ist das harmonische Mittel aussagekräftiger.

Das **harmonische Mittel** $\bar{x}_h$ ist definiert als der reziproke Wert des arithmetischen Mittels der reziproken Werte der einzelnen Merkmalsausprägungen:

$$\bar{x}_h = \frac{1}{\frac{1}{n} \cdot \left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n} \right)}$$

Mit dieser Formel ist das harmonische Mittel $\bar{x}_h$ zunächst nur für von Null verschiedene Zahlen x_i definiert. Geht aber einer der Werte x_i gegen null, so existiert der Grenzwert des harmonischen Mittels und ist ebenfalls gleich null. Daher ist es sinnvoll, das harmonische Mittel als Null zu definieren, wenn mindestens eine der zu mittelnden Größen gleich null ist.

Das harmonische Mittel $\bar{x}_h$ ist immer kleiner gleich dem geometrischen Mittel $\bar{x}_g$, und dieses war kleiner gleich dem arithmetischen Mittel $\bar{x}$. Falls wir das Minimum aller Merkmalsausprägungen $\min(x_1, x_2, x_3, \dots, x_n)$ und das Maximum aller Merkmalsausprägungen $\max(x_1, x_2, x_3, \dots, x_n)$ betrachten, ergibt sich die Ungleichung:

$$\min(x_1, x_2, x_3, \dots, x_n) \leq \bar{x}_h \leq \bar{x}_g \leq \bar{x} \leq \max(x_1, x_2, x_3, \dots, x_n)$$

Im Falle von $n = 2$ lässt sich das harmonische Mittel indirekt berechnen als:

$$\bar{x}_h = \frac{\bar{x}_g^2}{\bar{x}}$$

Beispiel 2.10

In den vier Abteilungen eines Kaufhauses werden die in der Tab. 2.5 angegebenen monatlichen Umsätze erzielt.

Schlüsselkonzepte zur Statistik
die wichtigsten Methoden, Verteilungen, Tests
anschaulich erklärt

Benesch, Th.

2013, X, 240 S. 68 Abb., Softcover

ISBN: 978-3-8274-2771-7