Kapitel 4: Mittelwerte und andere Lagemaße

1. Eigenschaften von Mittelwerten

Verteilungsmaßzahlen sollen bestimmte Charakteristika einer Häufigkeitsverteilung hervorheben und durch eine Zahl kennzeichnen. Im Falle der Mittelwerte gilt es das Niveau zu charakterisieren, also die allgemeine Größenordnung der Messwerte, bzw. die "zentrale Tendenz" (Lage des Zentrums) einer Häufigkeitsverteilung durch Angabe eines "typischen Werts" (Stellvertreter-Bedeutung des Mittelwerts). Mittelwerte sind bei weitem die bekanntesten statistischen Größen, deren Aussage allgemeinverständlich ist: ein Mittelwert ist, wie schon der Name sagt, ein mittlerer Wert, der einen Datensatz durch eine einzige Zahl "zusammenfasst".

Mittelwerte sind nicht die einzigen, wohl aber die bekanntesten und elementarsten Verteilungsparameter. Andere Aspekte einer Häufigkeitsverteilung sind Streuung, Schiefe, Wölbung, Konzentration und Disparität. Zu den entsprechenden Verteilungsmaßzahlen vgl. Kap. 5 und 6. Diese Maßzahlen sollten bei der Interpretation eines Mittelwerts mitberücksichtigt werden. So ist z.B. ein Mittelwert, isoliert betrachtet, immer dann nicht sehr aussagefähig, wenn die Einzelwerte sehr unterschiedlich sind, d.h. die Streuung groß ist.

<u>Def. 4.1: Mittelwertaxiome</u>

Mittelwerte M sind Verteilungsmaßzahlen, die unter Berücksichtigung des Skalenniveaus die folgenden Axiome M1 bis M5 erfüllen:

M1 **Einschrankung**: Es gilt bei der Größe nach geordneten Einzelwerten $x_{(1)} \leq M \leq x_{(n)}$ bzw. bei Merkmalsausprägungen $x_1 \leq M \leq x_m$.

M2 **Ergänzung**: Tritt zu den n Beobachtungswerten $x_1,x_2,...,x_n$ mit dem Mittelwert $M(x_1,...,x_n) = M_n$ ein weiterer Wert x_{n+1} hinzu, so soll für den "neuen" Mittelwert $M(x_1,...,x_{n+1}) = M_{n+1}$ gelten:
wenn $x_{n+1} \leq M_n$ dann $M_{n+1} \leq M_n$
wenn $x_{n+1} \geq M_n$ dann $M_{n+1} \geq M_n$

M3 **Transformation**: Für den Mittelwert M^* der transformierten Beobachtungswerte $x_v^* = f(x_v)$ soll gelten: $M^* = f(M)$. Dabei ist f eine auf dem Skalenniveau des Merkmals X zulässige Transformation.

M4 **Monotonie**: Bei den Merkmalen X und Y mit den Beobachtungsvektoren (Vektoren der Beobachtungswerte) **x** und **y** soll die Mittelwertfunktion monoton zunehmen in bezug auf die Beobachtungswerte bzw. Merkmalsausprägungen.
Für $\mathbf{x} \geq \mathbf{y}$ (vgl. Bem. Nr. 4) gilt $M(\mathbf{x}) \geq M(\mathbf{y})$.

M5 **Unabhängigkeit von den absoluten Häufigkeiten**: Für ein reelles k und mit den Vektoren **x** der Merkmalsausprägungen und **n** der absoluten Häufigkeiten gilt $M(\mathbf{x},\mathbf{n}) = M(\mathbf{x},k\cdot\mathbf{n})$ (d.h. eine Ver-k-fachung der absoluten Häufigkeiten verändert den Mittelwert nicht).

<u>Bemerkungen zu Def. 4.1:</u>

1. Im Axiom M1 wird von einem Mittelwert gefordert, dass er zwischen dem kleinsten und dem größten Beobachtungswert (einschließlich) liegt. Das entspricht auch dem umgangssprachlichen Verständnis des Wortes "Mittel"- Wert. Aus M1 folgt, dass für $x_v = c$ (für alle v =1,...,n) M = c ist. Sind alle Beobachtungen gleich (d.h. gleich der Konstanten c), dann ist auch der Mittelwert gleich dieser Konstanten c und auch ganz offensichtlich der charakteristische Wert.

2. Die in M2 gestellte Forderung gewinnt an Bedeutung bei der Untersuchung der "Robustheit" einer Maßzahl. Ein Hilfsmittel hierzu kann die sog. Einflusskurve (influence curve, Inf(...)) sein:

   <u>Def. 4.1a:</u>

   $\text{Inf}(M_n,x_{n+1}) = (n+1)(M_{n+1} - M_n)$ [Einflusskurve]
   welche die Abhängigkeit der (n+1)-fachen Mittelwertdifferenz bei Hinzukommen einer (n+1)-ten Beobachtung (d.h. dem Merkmalswert x_{n+1}) von diesem Wert x_{n+1} beschreibt. Bei Darstellung der einzelnen Mittelwerte wird auf dieses Konzept zurückgegriffen.

3. Eine häufig betrachtete Transformation ist die Lineartransformation (Gl. 4.1). Axiom M3 bedeutet dann, dass aus

(4.1)	$x_v^* = a + bx_v$	(b>0) folgt
(4.2)	$M^* = a + bM$,	

was eine problematische Forderung ist. Dies generell zu fordern [oder auch für andere, nichtlineare Transformationen], wie es P.J. Bickel und E.L. Lehmann[1] tun, halten wir nicht für sinnvoll, da Gl. 4.1 eine höchstens auf dem Intervallskalenniveau zulässige Transformation ist. Gl. 4.2 wird vom arithmetischen, nicht aber vom geometrischen und harmonischen Mittel erfüllt.

Andererseits wäre es nicht sinnvoll, wenn ein Mittelwert translationsinvariant (a hätte keinen Einfluss auf M^*) oder skaleninvariant (b hätte keinen Einfluß auf M^*) wäre.

In einer weiteren Bedingung fordern Bickel und Lehmann übrigens (4.3) $M(-\mathbf{x}) = -M(\mathbf{x})$ [entspricht a = 0 und b = -1 in Gl. 4.2], d.h. ändern alle Beobachtungswerte ihr Vorzeichen, so soll auch der Mittelwert sein Vorzeichen ändern. Als Mittelwerte sollten jedoch auch solche Kennzahlen zugelassen werden, die nur sinnvoll bei positiven Merkmalswerten zu berechnen sind, wie z.B. das geometrische Mittel.

4. Unter $\mathbf{x} \geq \mathbf{y}$ verstehen wir, dass $x_v \geq y_v$ (v = 1,...,n) ist, wobei für mindestens ein v $x_v > y_v$ gilt. Für die im Abschnitt b) dargestellten Mittelwerte gilt strenge Monotonie, d.h. es gilt unter den genannten Voraussetzungen $M(\mathbf{x}) > M(\mathbf{y})$. Dies sind Mittelwerte, die in bezug auf Unterschiedlichkeit der Merkmalswerte reagibler sind. Da ein Lokalisationsparameter das Niveau der Messwerte widerspiegeln soll, wäre es nicht sinnvoll, Mittelwerte zuzulassen, für die M4 nicht gilt. Denn das würde bedeuten, dass obgleich kein x-Wert kleiner als ein korrespondierender y-Wert ist, der Mittelwert der x-Werte kleiner als der Mittelwert der y-Werte sein kann.

   Für den Fall, dass **y** aus **x** durch eine zulässige Transformation hervorgeht, ist mit M3 auch M4 erfüllt.

5. Mit M5 wird sichergestellt, dass ein Mittelwert nicht abhängig ist von der Anzahl der Beobachtungen. Es ist daher unerheblich, ob ein Mittelwert mit absoluten oder mit relativen Häufigkeiten berechnet wird.

6. Es mag plausibel erscheinen, dass im Falle einer symmetrischen Verteilung ein "Mittelwert" im "Zentrum der Symmetrie" liegt. Obgleich diese Forderung vom arithmetischen Mittel, Median (Zentralwert) und Modus erfüllt wird, ist es nicht sinnvoll, diese Eigenschaft generell zu fordern. Denn dann - sofern nicht alle Beob-

---

1 Descriptive Statistics for Nonparametric Models, Part II: Location, in: The Annals of Statistics Vol 3 (1975), S.1045 - 1069.

achtungswerte identisch sind - liegen das geometrische und harmonische Mittel links vom Zentrum.

2. Spezielle Mittelwerte für metrisch skalierte Merkmale

In diesem und im folgenden Abschnitt werden die gebräuchlichsten Mittelwerte und ihre Eigenschaften vorgestellt. Das arithmetische, geometrische und harmonische Mittel erfordern eine metrische Skala, der Median mindestens eine Ordinalskala und der Modus nur eine Nominalskala. Einige weitere Mittelwerte werden im Zusammenhang mit speziellen Fragestellungen in anderen Kapiteln behandelt[2]. Wie gezeigt werden kann, erfüllen alle in diesem Abschnitt behandelten Mittelwerte die Axiome M1 bis M5.

a) Das arithmetische Mittel

Def. 4.2: arithmetisches Mittel

Die Maßzahl

(4.4)	$\bar{x} = \frac{1}{n}\sum x_\nu$	(Berechnung aus Einzelbeobachtungen) [**ungewogenes arithmetisches Mittel**]

oder

(4.5)	$\bar{x} = \frac{1}{n}\sum x_i n_i$	(Berechnung aus Merkmalsausprägungen)
	$= \sum x_i h_i$	[**gewogenes arithmetisches Mittel**]

heißt arithmetisches Mittel.

Hinweis:
Die Namen "gewogenes" und "ungewogenes" arithmetisches Mittel sollten nicht den Eindruck entstehen lassen, dass es sich um zwei verschiedene Mittelwerte handelt. Es sind nur Bezeichnungen für zwei Arten der Berechnung des gleichen arithmetischen Mittels, je nachdem, in welcher Form die Daten gegeben sind: einmal als Einzelbeobachtungen, zum anderen als gruppierte Daten (d.h. wenn Daten als Häufigkeitsverteilung vorliegen). Der Fall klassierter Daten wird im folgenden behandelt unter Nr.6.

2 Etwa der schwerste Wert und der Scheidewert im Kapitel 6.

Eigenschaften und Interpretation des arithmetischen Mittels:

1. Man kann leicht zeigen, dass $\bar{x}$ die Axiome M1 bis M5 erfüllt.
2. Schwerpunkt- oder Ausgleichseigenschaft

Satz 4.1: Schwerpunkteigenschaft

Es gilt

a) $\sum x_v - \bar{x} = 0$ bzw.

b) $\sum(x_i - \bar{x})h_i = 0$.

Beweis

a) $\Sigma(x_v - \bar{x}) = \Sigma x_v - n\,\bar{x} = 0$ da wegen (4.4) $\Sigma x_v = n\bar{x}$ gilt.

b) ist analog zu zeigen (es ist auch $\Sigma(x_i - \bar{x})\,n_i = 0$).

Beispiel 4.1 macht deutlich, warum hier vom "Schwerpunkt" gesprochen wird (vgl. Abb. 4.1). Die Summen der negativen (S_1) und positiven Abweichungen (S_2) der Merkmalswerte vom arithmetischen Mittel sind absolut gleich, nämlich $S_1 = -27$ und $S_2 = +27$.
Der **Schwerpunkt** ist bekanntlich der Quotient aus der Summe der "Momente", d.h. der Produkte aus Gewichten n_i und Hebelarmen x_i, also $\Sigma x_i\, n_i$ und der Summe der Gewichte Σn_i. In Abb. 4.1 erscheinen die Größen n_i als Gewichte (Symbol •) an einer Balkenwaage. Wird diese im Punkt $\bar{x} = 7$ unterstützt, so befindet sie sich im Gleichgewicht. Die Gleichheit von S_1 und S_2 bedeutet, dass sich negative und positive Abweichungen ausgleichen (**Ausgleichseigenschaft** des arithmetischen Mittels): Angenommen die Beträge 0,5,10,15 und 20 seien Einkommensbeträge (z.B. in 100 DM), dann wäre 7 (also 700 DM) genau der Betrag, den alle verdienen würden, wenn der Gesamtverdienst (die Summe aller Verdienste) auf alle n Einheiten gleich verteilt würde, d.h. wenn alle gleich viel verdienen würden, und S_1 und S_2 wären die Summen der unter-, bzw. der überdurchschnittlichen Einkommen.
Wird die **Merkmalssumme** $S = \Sigma\, x_i\, n_i$ auf n Einheiten (im Bsp. 4.1 ist $S = 70$ und $n = 10$) zu gleichen Teilen (jeweils 1/n-tel von S) verteilt, so erhält jede Einheit den Betrag $\bar{x} = S/n$. Das ist die Ausgleichs- oder Ersatzwerteigenschaft des arithmetischen Mittels. Folgerung:

> Das arithmetische Mittel bleibt unverändert bei solchen "Umverteilungen" zwischen den Merkmalswerten, bei denen die Merkmals**summe** konstant bleibt.

3. Minimumeigenschaft

Satz 4.2:

Die Funktion $Q(M) = \sum (x_v - M)^2$ besitzt ein Minimum an der Stelle $M = \bar{x}$, d.h. für alle $M \neq x,^{-}$ ist: $\sum (x_v - M)^2 > \sum (x_v - \bar{x})^2$.

Beweis:

$$\text{Aus} \quad \frac{dQ(M)}{dM} = 2 \sum (x_v - M)(-1) = 0$$

folgt $\Sigma x_v - nM = 0$ und somit $M = (\sum x_v)/n = \bar{x}$.
Da die zweite Ableitung nach M positiv ist, nämlich $dQ^2(M)/dM^2 = 2n > 0$, ist die Behauptung bewiesen. Die Summe der quadrierten Abweichungen hat kein Maximum [sie kann also beliebig groß sein], sondern nur ein Minimum.

Anmerkungen zum Beweis:

1. Der Beweis lässt sich auch mit dem Steinerschen Verschiebungssatz $\Sigma (x_v\text{-}a)^2 = \Sigma (x_v\text{-}\bar{x}) + n(\bar{x}\text{- } a)^2$ führen [Kap.5].
2. Wie man aus der notwendigen Bedingung $-2\ \Sigma\ (x_v - M) = 0$ sieht, folgt die Schwerpunkteigenschaft aus der Minimumeigenschaft.

Interpretation der Minimumeigenschaft:

Deutet man x_v als Summe einer (systematischen) Niveaukomponente m, die additiv überlagert ist von einer Fehlerkomponente u_v, also $x_v = m + u_v$, so ist $\bar{x} = m$ derjenige Wert der Niveaukomponente, um den die Abweichungen bzw. Fehler $u_v = x_v - m$ geringst möglich streuen (vgl. Varianz, Kapitel 5).

4. Lineartransformation, Voraussetzungen hinsichtlich der Skalen

Satz 4.3:

Das arithmetische Mittel erfüllt das Mittelwertaxiom M3 für lineare Transformationen. Aus Gl. 4.1 folgt in Verbindung mit Def. 4.2 für das arithmetische Mittel der transformierten Werte
$x_v^* = a + b\, x_v$:

$$\boxed{(4.6) \quad \bar{x}^* = a + b\bar{x} \qquad \text{(a,b reelle Zahlen).}}$$

Der Beweis ist elementar.
Damit ist auch gezeigt, dass das arithmetische Mittel bei Merkmalen, die auf einer Intervallskala [die ja invariant ist gegenüber linearen Transformationen] gemessen sind, anwendbar ist.

Sofern $b > 0$ ist, hat das Merkmal X^* die selben Skaleneigenschaften wie X. Somit ist die Lineartransformation eine "zulässige" Transformation.
Eine häufig benutzte Transformation ist die **Zentrierung**, d.h. die Bildung von Abweichungen vom arithmetischen Mittel, "zentrierter" Werte (sog. "deviation scores"):

$$z_v = x_v - \bar{x} \qquad (a = -\bar{x}\,,\ b = 1),$$

deren arithmetisches Mittel wegen Satz 4.3 Null ist.

5. Aggregationseigenschaft

Der folgende Satz läßt sich verallgemeinern für alle Mittelwerte, die ein Spezialfall des Potenzmittels sind.

Satz 4.4:

Wird eine Gesamtmasse (Gesamtheit) vom Umfang n zerlegt in g disjunkte Teilmassen (Teilgesamtheiten) mit den Umfängen n_1, $n_2,\ldots,n_g$, so dass $n = n_1 + n_2 + \ldots + n_g$, dann ist das arithmetische Mittel
mit

$$(4.7) \qquad \bar{x} = \sum_{j=1}^{g} \bar{x}_j h_j \qquad (j = 1,2,\ldots,g)$$

gegeben, wobei $\bar{x}_j$ das arithmetische Mittel der j-ten Teilgesamtheit ist:

$$(4.8) \qquad \bar{x}_j = \frac{1}{n_j} \sum_k x_{jk}\ . \qquad (k = 1,2,\ldots,n_j),$$

Dabei ist $h_j = n_j / n$ der Anteil der j-ten Teilgesamtheit am Gesamtumfang n, so dass der Gesamtmittelwert ein gewogenes arithmetisches Mittel der Teilmittelwerte ist.

Beweis

Mit S_j als Merkmalssumme der j-ten Teilgesamtheit $S_j = \sum_k x_{jk} = n_j \bar{x}_j$

ist die Merkmalssumme S der Gesamtmasse mit $S = \sum_j S_j = \sum_j \sum_k x_{jk}$

bzw. unter Berücksichtigung von Gl. 4.8 durch $S = \sum_j n_j \bar{x}_j$ gegeben.

Wegen $\bar{x} = S/n$ folgt hieraus unmittelbar Gl. 4.7.

6. Klassierte Daten

Sofern die **Klassenmittelwerte** $\bar{x}_k$ (k = 1,2,...,p) bekannt sind, berechnet man den Gesamtmittelwert $\bar{x}$ gem. Gl. 4.9:

$$(4.9) \quad \bar{x} = \sum_k \bar{x}_k h_k \qquad 1 \leq k \leq p \;.$$

Andernfalls verwendet man die **Klassenmitten** m_k (vgl. Kap. 3, Def. 3.5 b) und erhält den **geschätzten** Gesamtmittelwert m (als Schätzung von $\bar{x}$) mit:

$$(4.10) \quad m = \hat{x} = \sum_k m_k h_k$$

Im allgemeinen wird m von $\bar{x}$ verschieden sein. Die Näherung wird um so besser sein, je mehr sich die Beobachtungswerte (symmetrisch) um die Klassenmitten m_k verteilen.

Klassierte Daten stellen eine spezielle Form der in Satz 4.4 beschriebenen Zerlegung dar. Die Teilgesamtheiten sind aufgrund von aneinander angrenzenden Intervallen auf der x-Achse definiert (es ist dann g = p). Teilmassen können auch aufgrund anderer Kriterien (statt des x-Wertes) gebildet werden, wie z.B. nach dem Merkmal Geschlecht, Religion usw. Vgl. Beispiel 4.2 für eine Demonstration der Schätzung eines arithmetischen Mittels aus einer klassierten Verteilung.

Beispiel 4.1:

Man berechne das arithmetische Mittel für:

a) die folgenden 10 Merkmalswerte 0,5,0,10,15,5,0,20,10,5;
b) die folgende Häufigkeitsverteilung:

i	x_i	n_i
1	0	3
2	5	3
3	10	2
4	15	1
5	20	1

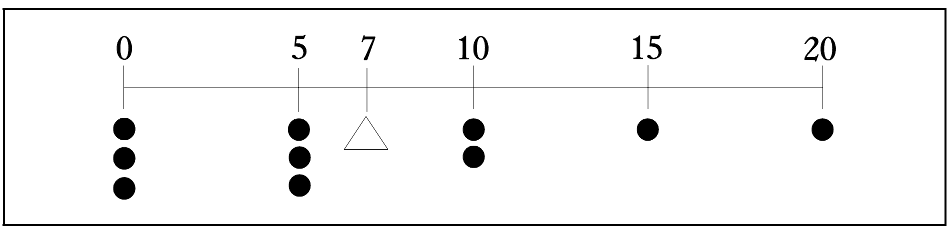
Lösung 4.1:

Es ist unschwer zu erkennen, dass es sich in beiden Fällen um die gleichen Daten handelt. Für das ungewogene arithmetische Mittel erhält man (0+5+0+10+15+5+0+20+10+5)/10 = 70/10 = 7 und für das gewogene arithmetische Mittel natürlich den gleichen Wert 7. Man berechnet das gewogene arithmetische Mittel am besten aus einer Arbeitstabelle, in der man in einer Spalte die Größen $x_i n_i$ bestimmt, deren Summe dann der Zähler von $\bar{x}$ ist (vgl. Abb. 4.1).

Abb. 4.1: Erläuterung des Begriffs "Gewicht" und der Schwerpunkteigenschaft von $\bar{x}$

i	x_i	n_i	$x_i n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})n_i$
1	0	3	0	-7	-21
2	5	3	15	-2	-6
3	10	2	20	+3	+6
4	15	1	15	+8	+8
5	20	1	20	+13	+13
S	-	10	70	-	0

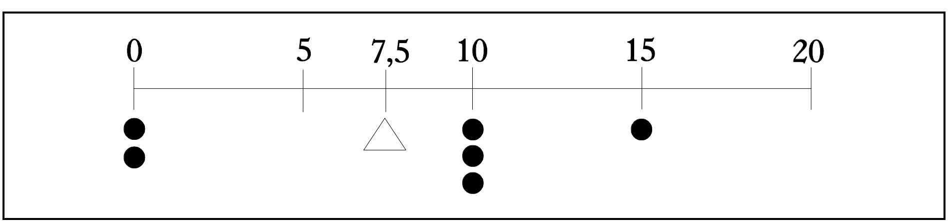
$\bar{x} = 70/10 = 7$



Variante von Bsp. 4.1: vgl. folgende Abbildung der Balkenwaage

i	x_i	n_i	$x_i n_i$
1	0	2	0
2	10	3	30
3	15	1	15

$\bar{x} = 45/6 = 7{,}5$ Im Punkt 7,5 befindet sich die Balkenwaage im Gleichgewicht.



Beispiel 4.2:

Berechnen Sie das arithmetische Mittel der Bruttomonatsverdienste von 150 Beschäftigten einer Unternehmung gemäß der folgenden Tabelle:

Klasse Nr. k	Bruttomonats-verdienst	Anzahl n_k
1	600 - 1000	4
2	1000 - 1400	11
3	1400 - 1800	19
4	1800 - 2200	31

Klasse Nr. k	Bruttomonats-verdienst	Anzahl n_k
5	2200 - 2600	37
6	2600 - 3000	22
7	3000 - 3400	18
8	3400 - 3800	8

Lösung 4.2:

Da die Klassenmittelwerte nicht bekannt sind, wird hier über die Klassenmitten m_k der Gesamtmittelwert m nach der Formel $m = \sum m_k h_k$ berechnet.

Ergebnis: m = 2305,60 DM (= durchschnittl. Bruttomonatsverdienst).

k	m_k	h_k	$m_k\, h_k$
1	800	0,027	21,6
2	1200	0,073	87,6
3	1600	0,127	203,2
4	2000	0,207	414,0
5	2400	0,247	592,8
6	2800	0,147	411,6
7	3200	0,120	384,0
8	3600	0,053	190,8
S			2305,6

Mittelwerte und Interpolation:

Das gewogene arithmetische Mittel von zwei Zahlen x_1 und x_2 (oder y_1 und y_2) stellt eine lineare Interpolation dar (Abb. 4.2).

Der Punkt $P(\bar{x},\bar{y})$ teilt die Strecke P_1P_2 zwischen den Punkten $P_1(x_1,y_1)$ und $P_2(x_2,y_2)$ im Verhältnis n_2/n_1 auf (vgl. Abb. 4.2, Teil a), denn

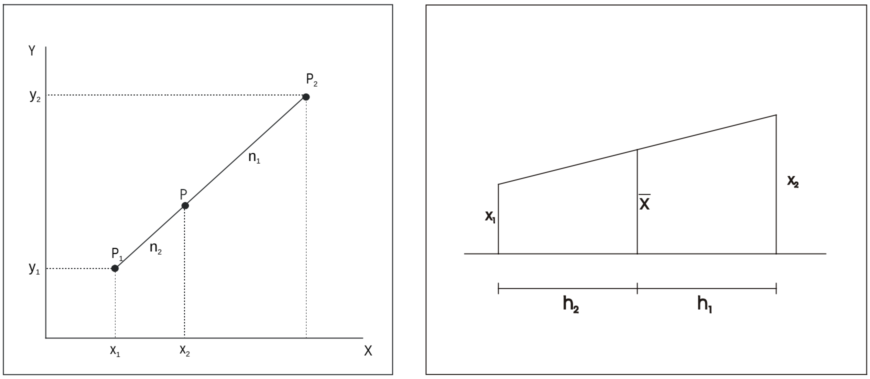
$$\bar{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2} \qquad \text{und} \qquad \bar{y} = \frac{n_1 y_1 + n_2 y_2}{n_1 + n_2}$$

Abb. 4.2, Teil b zeigt: $\bar{x}$ ist die lineare Interpolation zwischen den Höhen x_1 und x_2, denn $(x_2-x_1)/(h_1+h_2) = (\bar{x}-x_1)/h_2$ und $h_1 + h_2 = 1$.

Abb. 4.2: Lineare Interpolation

a) eine zweidimensionale Betrachtung (Variablen X,Y)

b) eindimensionale Betrachtung (h_1, h_2 als relative Häufigkeit)



"Mittelwerte der Pythagoräer"

Das ungewogene arithmetische Mittel zweier Zahlen x_1 , x_2 basiert auf der folgenden Forderung bezüglich des Verhältnisses zweier Abstände

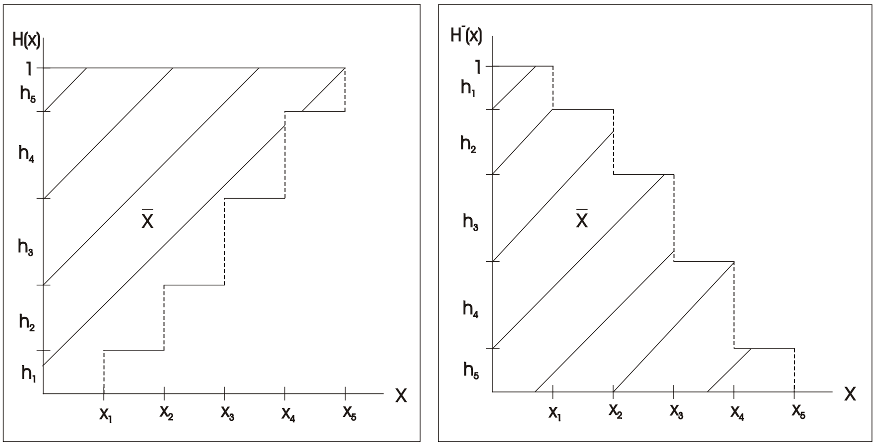
$$\frac{x_2 - M}{M - x_1} = \frac{x_2}{x_2} = \frac{x_1 - M}{M - x_2} = \frac{x_1}{x_1} = 1 \quad \Rightarrow \quad M = \overline{x} = \frac{x_1 + x_2}{2}$$

d.h. durch $\overline{x}$ wird die zwischen x_2 und x_1 gebildete Strecke genau halbiert.
Mit entsprechenden Forderungen an Verhältnisse von Abständen (Strecken) wurden in der Antike die sog. "Mittelwerte der Pythagoräer", nämlich $\overline{x}$, $\overline{x}_G$, $\overline{x}_H$ und $\overline{x}_A$ hergeleitet (vgl. Bemerkungen zu den entsprechenden Mittelwerten).

arithmetisches Mittel und Summenhäufigkeitsfunktion:

Es ist leicht zu sehen, dass bei einem nichtnegativen Merkmal X das arithmetische Mittel die Fläche zwischen der Ordinate (Häufigkeitsachse) und der Summenhäufigkeitskurve H(x) ist (d.h. die schraffierte Fläche in Abb. 4.3), bzw. die Fläche unterhalb der Resthäufigkeitskurve $H^-(x)$.

Abb.4.3: Das arithmetische Mittel als Fläche



b) Das geometrische Mittel

Das geometrische Mittel ist weit weniger bekannt und gebräuchlich als das arithmetische Mittel. Es wird vor allem im Zusammenhang mit Wachstumsfaktoren (vgl. Kap.9) benötigt.

<u>Def. 4.3: geometrisches Mittel</u>

Die Maßzahl

(4.11)	$\bar{x}_G = \left(\prod_{v=1}^{n} x_v\right)^{1/n}$	(bei Einzelbeobachtungen, "ungewogen"),

(das Produktzeichen P bedeutet Px_v = x_1 x_2...x_n), bzw.

(4.12)	$\bar{x}_G = \prod_{i=1}^{m} x_i^{h_i}$	(gruppierte Daten, "gewogen")

heißt geometrisches Mittel (der positiven Merkmalswerte x > 0).

<u>Beispiel 4.3:</u>

Man berechne das geometrische Mittel für

a) die folgenden Beobachtungen: 5,5,5,10,10,15,20,20;
b) die folgende Häufigkeitsverteilung:

i	x_i	n_i
1	5	3
2	10	2
3	15	1
4	20	2

$(n = \Sigma\, n_i = 8)$.

Lösung 4.3:

Es ist offensichtlich, dass es sich bei diesem Zahlenbeispiel in beiden Fällen um die gleichen Daten handelt. Man erhält:

für das ungewogene geometrische Mittel $(5 \cdot 5 \cdot 5 \cdot 10 \cdot 10 \cdot 15 \cdot 20 \cdot 20)^{1/8} = (75.000.000)^{1/8} = 9{,}64679$

für das gewogene geometrische Mittel natürlich den gleichen Wert, nämlich $(5^3 10^2 15^1 20^2)^{1/8} = 5^{3/8} 10^{2/8} 15^{1/8} 20^{2/8}$. Das arithmetische Mittel ist größer, nämlich 90/8 = 11,25.

Eigenschaften und Interpretation des geometrischen Mittels:

1. Das geometrische Mittel erfüllt die Axiome M1 bis M5.

2. Aus Def. 4.3 folgt unmittelbar

$$(4.13) \quad \log \bar{x}_G = \frac{1}{n} \sum \log(x_v) = \overline{\log(x)}$$

   und entsprechend bei gruppierten Daten, so dass der Logarithmus des geometrischen Mittels gleich ist dem arithmetischen Mittel der logarithmierten Merkmalswerte. Das geometrische Mittel wird deshalb auch logarithmisches Mittel genannt.

3. Die der Schwerpunkteigenschaft des arithmetischen Mittels entsprechende Eigenschaft des geometrischen Mittels ist die folgende Eigenschaft, die auch gelegentlich "Einseigenschaft" genannt wird, d.h. der Ausgleich relativer Größen (im Verhältnis zu $\bar{x}_G$). Es gilt:

   a) $\mathrm{P}(x_v / \bar{x}_G) = (\bar{x}_G)^{-n}\, \mathrm{P} x_v = (\bar{x}_G)^{-n} (\bar{x}_G)^n = 1$, bzw.
   b) $\mathrm{P}(x_i / \bar{x}_G)^{h_i} = 1$.

   Hierauf beruht die Eignung von $\bar{x}_G$ zur Mittelung von Wachstumsfaktoren (vgl. Kap. 9). Dies ist die wichtigste Anwendung des geometrischen Mittels in der Statistik. $\bar{x}_G$ hat für Vervielfachungsgrößen Bedeutung und immer dann Sinn, wenn das Produkt der Einzelwerte eine sinnvolle Größe ist.

4. Während $\overline{x}$ invariant ist gegenüber "Umverteilungen" dergestalt, dass die **Merkmalssumme** konstant ist, gilt dies für $\overline{x}_G$ hinsichtlich des **Produkts** der Merkmalswerte. Eine Verdoppelung (Erhöhung um 100%) von x_1 wird in diesem Sinne durch eine Halbierung (Senkung um 50%) von x_2 "ausgeglichen", so dass das ungewogene $\overline{x}_G$ unverändert bleibt.

   Man beachte:
   In diesem Beispiel hat sich x_1 um 100% erhöht und x_2 um 50% verringert. Ein Ausgleich findet mithin zwischen den Wachstumsfaktoren, nicht zwischen den Wachstumsraten statt. Nicht richtig wäre somit z.B., dass sich eine 10%ige Zunahme durch eine 10%ige Abnahme ausgleiche (Beispiel 4.4).

5. In Abb. 4.4 liegen alle Kombinationen von x_1 und x_2, die zu dem gleichen Wert des geometrischen Mittels $\overline{x}_G = \sqrt{x_1 x_2}$ führen auf der eingezeichneten Hyperbel. Entsprechend ist die Gerade AB der geometrische Ort aller Kombinationen x_1, x_2, die zum gleichen arithmetischen Mittel führen. Die Abb. 4.4 zeigt auch, dass stets gilt $\overline{x}_G \leq \overline{x}$, wobei $\overline{x}_G = \overline{x}$ genau dann gilt, wenn $x_1 = x_2$. Der gleiche Zusammenhang wird auch mit dem Höhensatz der Planimetrie gezeigt (Abb. 4.5) wonach $x_1 x_2 \leq \overline{x}^2$ ist.

6. Das geometrische Mittel kommt in der Ökonomie in Gestalt der Cobb-Douglas-Produktionsfunktion vor. Das Produktionsergebnis ist danach ein gewogenes geometrisches Mittel der Einsatzmengen der Produktionsfaktoren. Der allgemeineren CES-Funktion entspricht das Potenzmittel.

Abb. 4.4

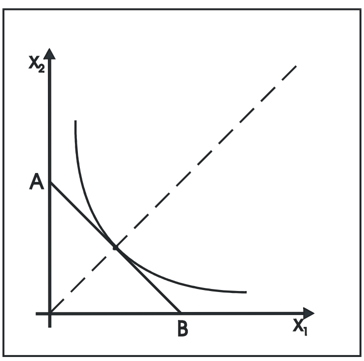
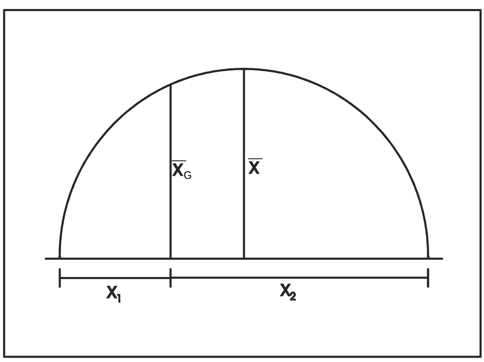


Abb. 4.5



7. Minimumeigenschaft

Satz 4.5:

Die Funktion $Q^*(M) = \sum[\log(x_v/M)]^2$ besitzt ein Minimum an der Stelle $M = \bar{x}_G$.

Beweis : Da $Q^*(M) = S\,[\log(x_v) - \log(M)]^2$ kann der Beweis analog Satz 4.2 geführt werden. Das Minimum ist an der Stelle $\log(M) = \overline{\log(x)}$, was nach Gl. 4.13 nichts anderes ist als $\log \bar{x}_G$. Es mag von Interesse sein, die Minimumeigenschaft des arithmetischen und des geometrischen Mittels an einem Beispiel zu vergleichen (vgl. Beispiel 4.5 und Abb. 4.6).

8. Deutet man x_v als ein Produkt einer (systematischen) Niveaukomponente m, die multiplikativ überlagert ist von einer Fehlerkomponente u_v, also $x_v = mu_v$, so ist $\bar{x}_G = m$ derjenige Wert der Niveaukomponente, um den die relativen Fehler $u_v = x_v/m$ geringst möglich streuen (vgl. Varianz, Kap. 5).

9. Zulässige Transformationen (erforderliche Skalen)

$\bar{x}_G$ erfüllt M3 für proportionale (linear-homogene) Transformationen. Aus $x_v^* = bx_v$ folgt

$$\bar{x}_G^* = \sqrt[n]{\prod x_v^*} = \sqrt[n]{\prod (bx_v)} = (b^n \prod x_v)^{1/n} = b\bar{x}_G.$$

Damit ist auch gezeigt, dass das geometrische Mittel nur bei Merkmalen, die mindestens auf einer Ratioskala gemessen sind, anwendbar ist.

10. Aggregation und klassierte Verteilung

Wird - wie in Satz 4.4 - eine Gesamtmasse (Gesamteinheit) vom Umfang n zerlegt in g disjunkte Teilmassen (Teilgesamtheiten), so ist $\bar{x}_G$ das mit den relativen Häufigkeiten h_j gewogene geometrische Mittel der geometrischen Mittel $\bar{x}_{Gj}$ der Teilgesamtheiten. Hieraus folgt für eine klassierte Verteilung:
Bei Kenntnis der geometrischen Mittel $\bar{x}_{Gk}$ (k = 1,2,...,p) der Klassen ist $\bar{x}_G$ als gewogenes geometrisches Mittel der (geometrischen) Klassenmittelwerte zu berechnen. Verwendet man statt dessen die Klassenmitten m_k, so kann $\bar{x}_G$ über- oder unterschätzt werden.

Die drei folgenden Eigenschaften sind für die Interpretation des geometrischen Mittels nicht sehr ergiebig, sie seien aber der Vollständigkeit halber aufgeführt:

11. Ersatzwerteigenschaft
    Ähnlich wie das arithmetische Mittel kann auch das geometrische Mittel $\bar{x}_G$ in einem bestimmten Sinn als Ersatzwert aufgefaßt werden. Die **Ersatzwert- oder Hochrechnungseigenschaft** von $\bar{x}_G$ bezieht sich auf das Gesamtmerkmalsprodukt Px_v mit $\bar{x}_G^n = Px_v$.

12. Das gewogene arithmetische Mittel zweier Zahlen x_1, x_2 stellt eine Interpolation gemäß einer Exponentialfunktion dar: $\bar{x}_G = x_1^{1-a}\ x_2^a = x_1(x_2/x_1)^a \quad 0 \leq a \leq 1$ mit a $= h_2$ im Unterschied zur linearen Interpolation $x_1 + a(x_2 - x_1)$ beim arithmetischen Mittel.

13. Das ungewogene geometrische Mittel zweier Zahlen x_1, x_2 basiert auf folgender Forderung bezüglich des Verhältnisses zweier Abstände:

$$\frac{x_2}{M} = \frac{M}{x_1} = \frac{x_2 - M}{M - x_1} \qquad \text{so dass } M = \bar{x}_G \text{ denn } \bar{x}_G / x_1 = x_2 / \bar{x}_G.$$

Beispiel 4.4:

Diplom-Kaufmann K aus E erhält im Jahre 1989 eine Gehaltserhöhung um 20%. Wegen der schlechten Geschäftslage im Jahre 1990 muss er jedoch 1990 eine Gehaltssenkung um 20% hinnehmen. Er verdient jetzt (Richtiges ankreuzen):

☐ weniger als ☐ mehr als ☐ genausoviel wie
vor der Gehaltserhöhung.

Lösung 4.4:

Nach einer einfachen Überlegung wird man „weniger“ ankreuzen, weil ja die Gehaltssenkung von 20% auf der Basis des gestiegenen Gehalts einen höheren Betrag ausmacht, als die vorhergehende Gehaltserhöhung. Es ergibt sich nömlich, wenn man einmal von DM 3000 ausgeht: Gehalt vor 1989: DM 3000, nach der Erhöhung: $3000 \cdot 1{,}2 = 3600$ DM und nach der Gehaltssenkung: $3600 \cdot 0{,}8 = 2880$ DM < 3000 DM.

Hinweis auf Kapitel 9:

Das Beispiel zeigt auch, warum es nicht sinnvoll ist, Wachstumsraten arithmetisch zu mitteln. Danach ergäbe sich als mittlere jährliche Wachstumsrate $x = ½[+20\% + (-20\%)] = 0\%$, was falsch ist, weil K ja nicht genauso viel verdient, wie vor der Gehaltserhöhung.

Der richtige Ansatz ist das geometrische Mittel der Wachstumsfaktoren $\sqrt{1{,}2\cdot 0{,}8}$ = $\sqrt{0{,}96}$ = 0,9798 = 1 - 0,0202. Hätte K jedes Jahr eine (konstante) Gehaltssenkung von 2,02% "erlitten", so wäre er zum gleichen Gehalt von DM 2880 gelangt, denn 3000·0,9798·0,9798 = 2880. Die mittlere jährliche Wachstumsrate beträgt also - 0,0202, d.h. -2,02%.

Beispiel 4.5:

Man bestimme und zeichne die Funktionen Q (gem. Satz 4.2) und Q^* (gem. Satz 4.5) für die folgenden Daten
$x_1 = 10$, $x_2 = 15$ und $x_3 = 20$
und für die folgenden Werte von M: 10,12,14,15,16,18,20.

Lösung 4.5:

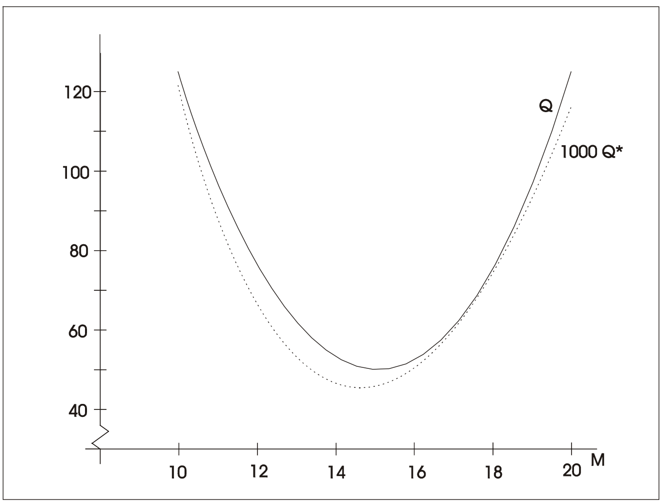
M	Q	1000 Q^*
10	125	121,6
12	77	64,9
14	53	46,2
14,42	51	45,7*)
15	50*)	46,6
16	53	51,8
18	77	73,5
20	125	106,2

*)Minimum

Die Mittelwerte sind $\bar{x} = 15$ und
$\bar{x}_G = 14{,}4224$.
Die genannten Funktionen lauten:
$Q = Q(M) = \Sigma\,(x_\nu - M)^2$
$Q^* = Q^*(M) = \Sigma\,[\log(x_\nu / M)]^2$

Man erkennt, dass die Funktion Q ihr Minimum an der Stelle M = 15 und die Funktion Q^* ihr Minimum an der Stelle M = 14,42 hat. Q ist symmetrisch um 15, Q^* ist aber nicht symmetrisch um 14,42.

Abb. 4.6



c) Das harmonische Mittel

Das harmonische Mittel hat in der Praxis eine ziemlich geringe Bedeutung. Es gibt jedoch Fälle, bei denen dieser weitgehend unbekannte Mittelwert angewendet werden *muss*, weil jede andere Art der Mittelwertbildung zu unsinnigen Ergebnissen führen würde. Das bekannteste Beispiel ist die Mittelung von Geschwindigkeiten, bzw. allgemeiner von Beziehungszahlen (Kap. 9). Das harmonische Mittel spielt auch eine Rolle bei den Indexzahlen (Kap. 10).

<u>Def. 4.4: harmonisches Mittel</u>

Die Maßzahl

(4.14) $\bar{x}_H = \dfrac{n}{\sum \dfrac{1}{x_v}}$ bei Einzelbeobachtungen, "ungewogen"

(4.15) $\bar{x}_H = \dfrac{n}{\sum \dfrac{n_i}{x_i}} = \dfrac{1}{\sum \dfrac{h_i}{x_i}}$ bei gruppierten Daten, "gewogen"

heisst harmonisches Mittel ($x \neq 0$).

<u>Beispiel 4.6: Mittelung von Geschwindigkeiten</u>

Ein Flugzeug legt für den Flug von A nach B und zurück insgesamt 4800 km zurück. Aufgrund von Gegenwind kann das Flugzeug auf dem Hinweg nur eine Geschwindigkeit von 600 km/h erreichen, auf dem Rückweg jedoch eine Geschwindigkeit von 800 km/h. Mit welcher Durchschnittsgeschwindigkeit ist das Flugzeug unterwegs?

<u>Lösung 4.6:</u>

Die Durchschnittsgeschwindigkeit wird mit dem harmonischen Mittel berechnet:

$$\bar{x}_H = \frac{n}{\sum \frac{n_i}{x_i}} = \frac{4800}{\frac{2400}{600} + \frac{2400}{800}} = \frac{2}{\frac{1}{600} + \frac{1}{800}} = 685{,}71$$

In diesem Fall ist das gewogene harmonische Mittel gleich dem ungewogenen, weil die beiden Strecken gleich lang sind.

Eine nach dem arithmetischen Mittel berechnete Durchschnittsgeschwindigkeit von 700 km/h wäre nicht richtig gewesen, denn das Flugzeug benötigt für die 2400 km des Hinflugs 4 Std. und für die 2400 km des Rückflugs 3 Std. Somit war das Flugzeug 7 Std. unterwegs um 4800 km zurückzulegen, woraus eine Durchschnittsgeschwindigkeit von 685,71 km/h folgt.

Eigenschaften und Interpretation des harmonischen Mittels:

1. Wie man leicht sieht, gilt:

der reziproke Wert von $\bar{x}_H$ ist das arithmetische Mittel der reziproken Werte (also der Werte $1/x_v$).

   Für die Berechnung ist es deshalb sinnvoll, zunächst den reziproken Wert $1/\bar{x}_H$ zu berechnen.

2. Man kann zeigen, dass $\bar{x}_H$ als Spezialfall des Potenzmittels alle Axiome M1 bis M5 erfüllt.

3. Das harmonische Mittel wird stets dann sinnvoll angewendet, wenn es gilt, Verhältniszahlen (siehe Kap. 9) zu mitteln, bei denen die im **Zähler** stehende Größe eine Konstante c ist und die im Nenner stehende Größe variabel ist. Das harmonische Mittel der Größen c/x_1, c/x_2,..., c/x_n beträgt nämlich $c/\bar{x}$, was sinnvoll interpretierbar ist. Entsprechend ist das arithmetische Mittel immer dann anzuwenden, wenn die im **Nenner** stehende Größe eine Konstante ist (vgl. Übers. 4.1).

Übersicht 4.1: Anwendung des arithmetischen und des harmonischen Mittels

Es ist anzuwenden das	
harmonische Mittel	**arithmetische Mittel**
bei einer Verhältniszahl) mit*	
$\dfrac{c}{x_v}$	$\dfrac{x_v}{c}$
Z konstant, N variabel	**Z variabel, N konstant**
bzw. bei einer Transformation von x_v in	
$y_v = \dfrac{c}{x_v}$	$z_v = \dfrac{x_v}{c} = \dfrac{1}{c} x_v$
(nichtlineare Transformation)	**(lineare Transformation)**

*) Eine Verhältniszahl (vgl. Kap. 9) ist ein Quotient mit einem Zähler Z und Nenner N. Die Größe c bezeichnet hier eine Konstante.

Zu welchen Ergebnissen eine arithmetische oder harmonische Mittelung führt sei an einem Beispiel mit n = 3 Beobachtungen gezeigt:

	Art der Mittelung	
	arithmetisch	harmonisch
y_v	$\bar{y} = (c/x_1 + c/x_2 + c/x_3)/3$	$\bar{y}_H = c/\bar{x}_H$ [sinnvoll]
z_v	$\bar{z} = \bar{x}/c$ [sinnvoll]	$\bar{z}_H = 3/(c/x_1 + c/x_2 + c/x_3)$

Man erkennt, dass die Größen $\bar{y}_H$ und $\bar{z}$ sinnvolle Größen sind, die Mittelwerte $\bar{z}_H$ und $\bar{y}$ dagegen keine sinnvollen Größen sind. Im speziellen Fall von n = 2 Einzelwerten gilt übrigens folgender Zusammenhang zwischen dem arithmetischen, harmonischen und geometrischen Mittel:

(4.16) $\bar{y} = ½(c/x_1 + c/x_2) \quad = \quad c\bar{x}/\bar{x}_{G,}^{2}$ und $\bar{z}_H = \bar{x}_{G,}^{2}/c\bar{x}$.

Drei Beispiele für die Anwendung des harmonischen Mittels

a) **Geschwindigkeiten**

Die Geschwindigkeit v auf einer gegebenen Strecke der Länge s ist umso größer (geringer), je kürzer (länger) die hierfür benötigte Zeit t ist, denn v = s/t. Die Durchschnittsgeschwindigkeit, die erreicht wird, wenn eine Strecke s zweimal befahren wird (hin und zurück) mit den Zeiten t_1 und t_2 beträgt dann

$$2s/(t_1 + t_2) = 2/(t_1/s + t_2/s) = 2/(1/v_1 + 1/v_2).$$

Das ist aber genau das harmonische Mittel der beiden Geschwindigkeiten $v_1=s/t_1$ und $v_2=s/t_2$. Diese Überlegung lässt sich für

a) n gleichlange Teilstrecken der Länge s
b) unterschiedlich lange Strecken $s_1,s_2,...,s_m$, bei denen die Geschwindigkeiten $v_1,v_2,...,v_m$ betragen,

verallgemeinern.

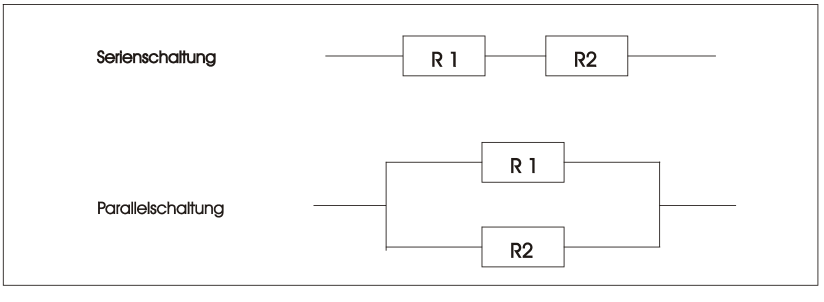
<u>zu a:</u> Ermittelt werden soll die Durchschnittsgeschwindigkeit v der Einzelgeschwindigkeiten v_v (v = 1,2,...,n) für die Gesamtstrecke ns. Hierzu kann man von den Identitäten t = ns/v und $t = \Sigma\, t_v = \Sigma\, (s/v_v)$ ausgehen. Aus beiden Gleichungen folgt $v = ns/(\Sigma\, s/v_v) = n/\Sigma\, (1/v_v)$, was nichts anderes ist, als das ungewogene harmonische Mittel der n Einzelgeschwindigkeiten v_v.

zu b: In diesem Fall ist $t = \Sigma t_i = \Sigma (s_i/v_i)$ (da für jede der m Teilstrecken gilt $v_i = s_i /t_i$) und $v = \Sigma s_i /\Sigma (s_i /v_i)$, d.h. das mit den Strecken s_i gewogene harmonische Mittel der Geschwindigkeiten v_i.
Die Durchschnittsgeschwindigkeit ist somit gleich dem harmonischen Mittel der Einzelgeschwindigkeiten. Das arithmetische Mittel führt (abgesehen vom Fall lauter gleicher Einzelgeschwindigkeiten) zu einer überhöhten mittleren Geschwindigkeit.

b) **Preise und Ausgaben**
Geldausgaben y_i sind Produkte aus Mengen q_i und Preisen $p_i=y_i /q_i$. Werden Preise bei konstanten Mengen (Nenner) gemittelt, so ist das arithmetische Mittel anzuwenden, sind die Mengen (aufgrund von Substitutionen) dergestalt veränderlich, dass die Ausgaben y (Zähler) konstant sind, so ist das harmonische Mittel anzuwenden. Ausgaben und gegebenenfalls auch Mengen sind aggregierbar, so dass $Y = \Sigma y_i$ und $Q = \Sigma q_i$. Gesucht ist ein mittlerer Preis (P) für den gilt $Y = P\cdot Q$. Es ist nun zu unterscheiden:
1. gleiche Mengen: sei $q_i = q$ für alle $i = 1,...,n$ Waren, dann ist $P = Y/Q = \Sigma y_i /nq = q\Sigma p_i /nq = \Sigma p_i /n$ das ungewogene arithmetische Mittel der Preise;
2. gleiche Ausgaben: $y_i = y$, dann ist P das ungewogene harmonische Mittel der Preise $P = Y/Q = ny/\Sigma (y/p_i) = n/\Sigma (1/p_i)$.

c) **Widerstände** (vgl. Abb. 4.7)
Die Unterscheidung zwischen arithmetischer und harmonischer Mittelung tritt auch auf bei der Schaltung von Widerständen (für den Widerstand R gilt R = U/I mit U = Spannung, I = Stromstärke). Ein Widerstand, der den zwei Widerständen gleichwertig ist, ist der sog. "Ersatzwiderstand" R_E. Dabei gilt:
bei Serien- (Reihen-) Schaltung $R_E = R_1 + R_2$
bei Parallelschaltung $1/R_E = 1/R_1 + 1/R_2$
(entsprechende Zusammenhänge gelten bei mehr als zwei Widerständen).

Abb. 4.7: Schaltung von Widerständen



Weitere Eigenschaften des harmonischen Mittels:

1. Zulässige Transformationen, Skalen
   $\bar{x}_H$ erfüllt M3 für proportionale Transformationen: Aus $x_v^* = bx_v$ folgt mit $\bar{x}_H^* = n/S\ (1/x_v^*) = b\bar{x}_H$. Die Berechnung des harmonischen Mittels setzt somit eine Ratioskala voraus.

Die folgenden Eigenschaften seien nur kurz erwähnt:

2. Der Schwerpunkteigenschaft des arithmetischen Mittels entspricht beim harmonischen Mittel eine Ausgleichseigenschaft der reziproken Werte.

   Satz 4.6:

   Es gilt

   a) $\Sigma(1/x_v - 1/\bar{x}_H) = \Sigma(x_v^{-1} - \bar{x}_H^{-1}) = 0$ oder mit n multipliziert $\sum \frac{1}{x_v}(x_v - \bar{x}_M) = 0$ [3]

   b) $\Sigma\ (1/x_i - 1/\bar{x}_H)h_i = 0$

   Beweis

   a) $\Sigma\ x_v^{-1} - n/\bar{x}_H = 0$ da nach Definition $\bar{x}_H = n/\Sigma\ x_v^{-1}$ ist;

   b) gilt entsprechend.

3. Minimumeigenschaft: Die Funktion $\Sigma\ (c/x_v - c/M)^2$ mit der reellen Konstante c besitzt ein Minimum an der Stelle $M = \bar{x}_H$. Beweis: analog den Ausführungen beim arithmetischen Mittel.

4. Schwerpunkt- und Minimumeigenschaft kann man auch wie folgt darstellen. Es gilt:
   $\sum x_v^r (x_v - M) = 0$
   $\sum x_v^r (x_v - M)^2 = \text{Min}$ (bzgl. M)
   für folgende Mittelwerte:
   $M = \bar{x}_H$ (harmonisch): r = -1
   $M = \bar{x}$ (arithmetisch): r = 0
   $M = \bar{x}_A$ (antiharmonisch): r = 1 (Def. 4.5)

5. Aggregation und klassierte Daten
   Wird eine Gesamtmasse vom Umfang n zerlegt in g disjunkte Teilmassen mit den Umfängen $n_1, n_2, ..., n_g$ $(n = n_1 + n_2 + ... + n_g)$ so gilt:

   $$\boxed{(4.17)\quad \bar{x}_H = n(n_1/\bar{x}_{H1} + ... + n_g/\bar{x}_{Hg})^{-1}}$$

   mit den (harmonischen) Teilmittelwerten $\bar{x}_{H1}, \bar{x}_{H2}, ..., \bar{x}_{Hg}$. Dieser Zusammenhang ist entsprechend anwendbar bei klassierten Daten.

---

[3] Zu einer inhaltlichen Interpretation dieser Eigenschaft des harmonischen Mittels vgl. Neubauer, W., Statistische Methoden, Frankfurt/M. 1991, S.62.

6. Von einer harmonischen Reihe $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ spricht man, weil jedes Glied das harmonische Mittel der benachbarten Glieder ist. Diese Reihe ist übrigens ein interessantes Beispiel dafür, dass eine Reihe nicht notwendig konvergiert, wenn ihre Glieder gegen Null streben.

7. Ersatzwerteigenschaft: Die Summe der reziproken Merkmalswerte $\Sigma\ 1/x_v$ kann aus dem reziproken Wert des harmonischen Mittels $\bar{x}_H$ gemäß $\Sigma\ 1/x_v = n\bar{x}_H^{-1}$ bestimmt werden.

8. Das gewogene harmonische Mittel zweier Zahlen x_1 , x_2 ist darstellbar als Interpolation gemäß der Funktion $\bar{x}_H = \frac{x_1\, x_2}{x_1 a + x_2(1\text{-}a)}$ mit $0 \leq a = h_2 \leq 1$.

9. Das ungewogene harmonische Mittel zweier Zahlen x_1 , x_2 resultiert aus der Forderung $(x_2 - M)/(M - x_1) = x_2 /x_1$ oder $(x_1 - M)/(M - x_2) = x_1 /x_2$ oder $x_2 /(x_2 - M) = (M - x_1) / x_1$ für das Teilungsverhältnis einer Strecke $x_1 - x_2$ (weitere Erläuterungen hierzu vgl. Abb. 4.9). Denn dann ist $M = 2\, x_1\, x_2 /(x_1+x_2) = \bar{x}_H$.

Zu weiteren Ausführungen über die die Anwendung von $\bar{x}_H$ und Zusammenhänge zwischen $\bar{x}$ und $\bar{x}_H$ vgl. Kap. 9 (Aggregation von Verhältniszahlen) und Kap. 10 (Zeitumkehrprobe bei Indexzahlen).

d) Quadratisches und antiharmonisches Mittel

Von sehr geringer Bedeutung für die Praxis sind das quadratische Mittel $\bar{x}_Q$ und das antiharmonische Mittel $\bar{x}_A$, die deshalb hier nur kurz zusammen behandelt werden und sinnvoll nur bei positiven Merkmalswerten berechnet werden.

<u>Def. 4.5: quadratisches- und antiharmonisches Mittel</u>

a) Das quadratische Mittel wird aus Einzelwerten ("ungewogen") mit

$$\boxed{(4.18) \quad \bar{x}_Q = +\sqrt{\frac{\Sigma x_v^2}{n}}}$$

bzw. bei gruppierten Daten (Merkmalsausprägungen, "gewogen") mit

$$\boxed{(4.19) \quad \bar{x}_Q = \sqrt{\Sigma\, x_i^2\, h_i}}$$

berechnet.

b) Die Maßzahl

$$\boxed{(4.20) \quad \bar{x}_A = \frac{\bar{x}_Q^2}{\bar{x}}}$$

heißt antiharmonisches Mittel.

Eigenschaften und Interpretation:

1. Das arithmetische Mittel der transformierten Werte $x_v^* = (x_v / \sqrt{\bar{x}})^2$ ist das antiharmonische Mittel der Werte x_v. Da dies eine monotone Transformation ist, gelten die vom arithmetischen Mittel erfüllten Axiome M1, M2 und M4 auch für $\bar{x}_A$. Da $\bar{x}_Q$ und $\bar{x}$ nur von den relativen, nicht den absoluten Häufigkeiten abhängen, erfüllt $\bar{x}_A$ auch M5. Das quadratische Mittel $\bar{x}_Q$ läßt sich als Potenzmittel darstellen, so dass x_Q alle Axiome M1 bis M5 erfüllt.
2. Für die transformierten Beobachtungswerte $x_v^* = bx_v$ gilt $\bar{x} = b\bar{x}_Q$ sowie $\bar{x}_A^* = b\bar{x}_A$, so dass $\bar{x}_Q$ und $\bar{x}_A$ bei Merkmalen, die auf einer Ratioskala gemessen sind, anwendbar sind.
3. Das quadratische Mittel tritt in der Streuungsmessung auf: die Standardabweichung ist das quadratische Mittel der Abweichungen vom arithmetischen Mittel.
4. Das ungewogene quadratische Mittel kann als ($\sqrt{1/n}$)-fache euklidische Distanz zwischen den Ortsvektoren x_1 , x_2 ,...,x_n im n-dimensionalen orthogonalen Koordina-

   tensystem aufgefaßt werden. Bei nur zwei Beobachtungen x_1 , x_2 gilt für diesen Ab-

   stand $\sqrt{x_1^2 + x_2^2} = \bar{x}_Q\sqrt{2}$

Nochmals: Mittelwerte der Pythagoräer

Es wurde bereits an den jeweiligen Stellen auf die formale Beschreibung der drei schon den Pythagoräern bekannten Mittelwerte $\bar{x}$, $\bar{x}_G$ und $\bar{x}_H$ eingegangen. Danach gelten für die Teilung der Strecke $x_2 - x_1$ (wenn $x_2 > x_1$)[4] in die beiden Streckenabschnitte M-x_2 und x_1-M (vgl. auch Abb. 4.8) die in Übersicht 4.2 aufgeführten Beziehungen.

Übersicht 4.2:

Mittel		
arithmetisch	$\frac{x_2 - M}{M - x_1} = \frac{x_2}{x_2} = \frac{½(x_2-x_1)}{½(x_2-x_1)} = 1$	wenn $M = \bar{x}$
geometrisch	$\frac{x_2 - M}{M - x_1} = \frac{x_2}{M} \geq 1$	wenn $M = \bar{x}_G$
harmonisch	$\frac{x_2 - M}{M - x_1} = \frac{x_2}{x_1} > 1$	wenn $M = \bar{x}_H$

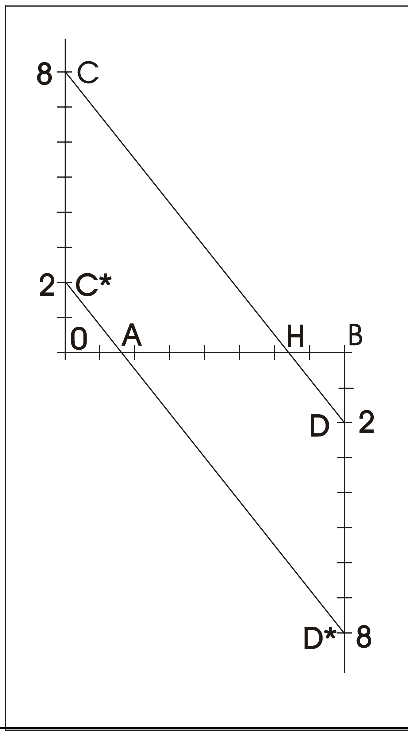
antiharmonisch	$\frac{x_2 - M}{M - x_1} = \frac{x_1}{x_2} < 1$	wenn $M = \overline{x}_A$,

Aus dieser Betrachtung wird auch die Bezeichnung "anti"-harmonisch verständlich.

Die geometrische Veranschaulichung dieser Forderungen soll mit einem einfachen Zahlenbeispiel in Abb. 4.8 erfolgen. Die Strecke OB ist die Differenz der beiden Einzelwerte $x_2 = 8$ und $x_1 = 2$. Es ist also $OB = x_2 - x_1 = 6$. Trägt man

- in O senkrecht nach oben die Strecke x_2 ($OC = x_2 = 8$) und in B senkrecht nach unten die Strecke x_1 ($BD = x_1 = 2$) an, so teilt die lineare Verbindung zwischen C und D die Strecke OB in die Abschnitte $OH = x_2 - \overline{x}_H = 8-3,2 = 4,8$ und $HB = \overline{x}_H - x_1 = 3,2-2 = 1,2$ (**harmonisches Mittel**).
- in O senkrecht nach oben die Strecke x_1 ($OC^* = x_1 = 2$) und in B senkrecht nach unten die Strecke x_2 ($BD^* = x_2 = 8$) an, so teilt die lineare Verbindung zwi-schen C^* und D^* OB in die Abschnitte $OA = x_2 - \overline{x}_A = 1,2$ und $AB = \overline{x}_A - x_1 = 4,8$ (**antiharmonisches Mittel**).

Abbildung 4.8: harmonisches und antiharmonisches Mittel



Das arithmetische Mittel (Punkt M) teilt OB in genau zwei Hälften. Beim geometrischen Mittel $\overline{x}_G$ (Punkt G [$\overline{x}_G = 4$, so dass $OG = x_2 - \overline{x}_G = 8-4 = 4$]) gilt genauso wie beim harmonischen Mittel $(x_2-M) / (M-x_1) > 1$ (weshalb die Punkte G und H rechts von M liegen), während beim antiharmonischen Mittel diese Relation < 1 ist (Punkt A liegt links von M).

[4] Die Formeln für $x_1 > x_2$ erhält man durch Vertauschung von x_1 und x_2.

e) Das Potenzmittel

Das sog. Potenzmittel ist eine Verallgemeinerung der in diesem Abschnitt behandelten Mittelwerte, so dass das harmonische, geometrische, arithmetische und quadratische Mittel jeweils Spezialfälle des Potenzmittels darstellen.

Def. 4.6: Potenzmittel

Die folgende Klasse von Mittelwerten

$$(4.21)\quad \bar{x}_{P,r} = \left[\frac{1}{n}(x^r_{,1} + x^r_{,2} + \ldots + x^r_{,n})\right]^{1/r} = \left(\frac{1}{n}\sum x^r_v\right)^{1/r}$$

(ungewogene Berechnung), bzw.

$$(4.22)\quad \bar{x}_{P,r} = (x_1^r h_1 + x_2^r h_2 + \ldots + x_m^r h_m)^{1/r} = \left(\sum x_i^r h_i\right)^{1/r}$$

(gewogene Berechnung)

heißt Potenzmittel.

Bemerkungen zu Def. 4.6:

1. Es ist unschwer zu erkennen, dass die folgenden Mittelwerte Spezialfälle des Potenzmittels sind:

   r = -1: harmonisches Mittel $\bar{x}_{p,-1} = \bar{x}_H$
   $r \to 0$: geometrisches Mittel $\bar{x}_{p,0} = \bar{x}_G$
   r = +1: arithmetisches Mittel $\bar{x}_{p,1} = \bar{x}$
   r = +2: quadratisches Mittel $\bar{x}_{p,2} = \bar{x}_Q$.

   Es ist klar, dass r=0 nicht unmittelbar auf Gl. 4.22 angewendet werden kann, weil dies einen unbestimmten Ausdruck ergeben würde, bzw. für $\log(\bar{x}_{P,r}) = \log(n^{-1}\Sigma\, x_v^r)/r$ den ebenfalls unbestimmten Ausdruck 0/0 liefern würde. Mit der l´Hospitalschen Regel (Differenzieren von Zähler [also $\log(n^{-1}\,\Sigma\, x_v^r)$] und Nenner [also r] dieses Ausdrucks nach r) lässt sich jedoch zeigen, dass $\log(\bar{x}_{P,0})$ der Logarithmus des geometrischen Mittels ist.

2. Man kann zeigen, dass $x_{p,r}$ mit wachsendem r monoton zunimmt, so dass die Ungleichung von Cauchy gilt:

$$(4.23)\quad \bar{x}_H \le \bar{x}_G \le \bar{x} \le \bar{x}_Q$$

Die Gleichheit gilt dann und nur dann, wenn alle Beobachtungen x_v (v=1,2,...,n) gleich sind. Dass $\bar{x}_{p,r}$ mit wachsendem r monoton zunimmt, ergibt sich aus folgender Überlegung: Setzt man $f(x) = x^r$, was ja "nur" eine monotone Transformation ist, so ist $f(\bar{x}_{p,r}) = (\bar{x}_{p,r})^r = (1/n)\ \Sigma f(x_v) = (1/n)\ \Sigma x_v^r$, was mit wachsendem Exponenten r wächst (wie auch $\bar{x}_{p,r}$).

3. Mittelwerte und Lageparameter für nicht notwendig metrisch skalierte Merkmale

a) Zentralwert (Median)

<u>Def. 4.7: Median</u>

Das Merkmal X sei mindestens ordinalskaliert. Dann gilt für den Zentralwert (Median) $Z = \tilde{x}_{0,5}$

a) Bei Einzelbeobachtungen

$$(4.24)\quad Z = \tilde{x}_{0,5} = \begin{cases} x_{((n+1)/2)} & \text{falls n ungerade} \\ \frac{1}{2}[x_{(n/2)} + x_{([n/2]+1)}] & \text{falls n gerade} \end{cases}$$

Der Median ist der Wert, der in einer der Größe nach geordneten Reihe $x_{(1)} \leq x_{(2)} \leq \ldots\ x_{(n)}$ in der Mitte, d.h. an der ½(n+1)- ten Stelle steht (bzw. die Interpolation zwischen dem n/2-ten Wert und dem darauf folgenden Wert an der Stelle n/2 + 1).

b) Bei gruppierten Werten (Häufigkeitsverteilung) gilt entsprechend für den Median

$$(4.25)\quad \tilde{x}_{0,5} = \begin{cases} x_i & \text{falls } H_{i-1} < 0{,}5 \text{ und } H_i > 0{,}5 \\ \frac{1}{2}(x_i + x_{i+1}) & \text{falls } H_i = 0{,}5 \end{cases}$$

c) bei klassierten Daten wird der Median aus der Summenhäufigkeitskurve bestimmt (zur Interpolation vgl. Gl. 4.26, Bem. Nr. 7).

<u>Bemerkungen zu Def. 4.7</u>

1. Allgemein ist der Median der Merkmalswert, der die Daten in genau zwei gleiche Teile teilt: mindestens 50% der Merkmalswerte sind kleiner oder gleich Z und mindestens 50% aller Merkmalswerte sind größer oder gleich Z. Falls n gerade ist, kommt bei Einzelwerten somit jeder Wert im Intervall von $x_{(n/2)}$ bis zum folgenden Wert $x_{([n/2]+1)}$ als

Kandidat für den Median in Betracht. Entsprechend erfüllt für $H_i = 0,5$ bei gruppierten Daten jeder Wert im Intervall (x_i, x_{i+1}) diese Bedingung. Um eine eindeutige Bestimmung von Z zu gewährleisten, setzt man häufig (wie hier) in beiden Fällen Z jeweils als Intervallmitte fest.

2. Häufig interessiert nicht oder nicht nur der Durchschnitt im Sinne des arithmetischen Mittels, sondern vielmehr die Orientierung an der Mitte einer Verteilung. So bedeutet "Mittelmäßigkeit" im allgemeinen, dass hinsichtlich eines bestimmten Kriteriums etwa gleichviele Individuen besser und schlechter sind. Ähnlich ist bei ökonomischen Größen wie Einkommen, Vermögen und auch z.T. bei Preisen die Mitte der Verteilung häufig aufschlußreicher, als das arithmetische Mittel, da dies bei asymmetrischen Verteilungen nicht unbedingt ein "repräsentativer" oder "typischer" Wert zu sein braucht.

3. Während das arithmetische Mittel sehr reagibel gegenüber Ausreißern ist, ist der Median sehr robust.

4. Minimumeigenschaft

   Der Median hat die folgende Eigenschaft, die jedoch hier nicht bewiesen werden soll (man findet den Beweis aber in einigen Lehrbüchern der Statistik):

Die Summe der absoluten Abweichungen des Medians von den Beobachtungen x_v ist minimal.

   Während das arithmetische Mittel die Summe der quadrierten Abweichungen $\Sigma(x_v - M)^2$ minimiert, besitzt der Median diese Eigenschaft hinsichtlich der Summe der absoluten Abweichungen, d.h. $\Sigma|x_v - \tilde{x}_{0,5}|$ ist minimal. Beide Eigenschaften haben auch bei der Konstruktion von Streuungsmaßen eine Bedeutung.

5. Transformation

   Der Median erfüllt M3 für alle streng monotonen Transformationen, welche die Reihenfolge der Merkmalswerte nicht ändern. Bei nicht streng monotonen Transformationen läßt sich der Median der transformierten Daten im allgemeinen nicht dadurch bestimmen, dass man auf den Median der Ursprungswerte dieselbe Transformation anwendet. Mithin setzt eine Anwendung des Medians ein mindestens ordinal skaliertes Merkmal voraus.

6. Aggregation

   Im allgemeinen läßt sich der Median einer Gesamtheit nicht aus den Medianen von Teilgesamtheiten bestimmen. Es ist daher erforderlich, aus allen geordneten Reihen der Teilgesamtheiten, eine **einzige** geordnete Reihe der aus den Teilgesamtheiten bestehenden Gesamtheit zu bilden.

7. Klassierte Daten

   Bei klassierten Daten erfüllt der Median $\tilde{x}_{0,5}$ näherungsweise die Bedingung $H(x \le \tilde{x}_{0,5}) = 0{,}5$. Ein Näherungswert für den Zentralwert wird durch Interpolation bestimmt. Man ermittelt hierzu zunächst die **Medianklasse** (d.h. die Klasse, in die der Median "fällt") k mit der Eigenschaft (mit x_r' = Klassen**obergrenze** der Klasse r):

$$\boxed{H_{k-1} \le 0{,}5 < H_k, \text{ so dass gilt } x_{k-1}' \le \tilde{x}_{0,5} < x_k' .}$$

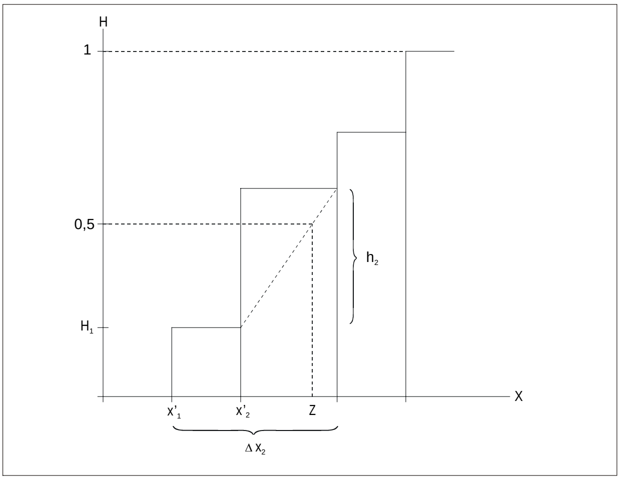
   Der Näherungswert für den Median lässt sich dann als Summe aus der Klassenuntergrenze x_{k-1}' der Medianklasse und einem Anteil der Klassenbreite b_k der k-ten Klasse (Medianklasse) definieren, der durch den Proportionalitätsfaktor $(0{,}5 - H_{k-1})/h_k$ gegeben ist (vgl. auch Abb. 4.9):
   Interpolation des Medians:

$$\boxed{(4.26) \quad \tilde{x}_{0,5} = x_{k-1}' + b_k(0{,}5 - H_{k-1})/h_k}$$

   Diese Formel ist unschwer aus einem "Dreisatz" zu entwickeln, denn offensichtlich gilt für den Betrag z, der zur Untergrenze x_{k-1}' der Medianklasse hinzuzuaddieren ist, gem. Abb. 4.9:
   $z/(\frac{1}{2} - H_{k-1}) = b_k/h_k$, so dass $\tilde{x}_{0,5} = x_{k-1}' + z$.

Der Einfachheit halber unterscheiden wir hier in der Notation nicht zwischen dem unbekannten Median und seinem Näherungswert. Mit Gl. 4.26 erhält man eine umso bessere Approximation, je mehr sich die Verteilung innerhalb der Medianklasse k einer Gleichverteilung (im Sinne gleicher Häufigkeiten aller Werte innerhalb der Klasse) nähert. Z läßt sich analog zu Gl. 4.26 auch aus den kumulierten absoluten Häufigkeiten N_k bestimmen.

Abb. 4.9: Bestimmung des Medians mit
Interpolation aus der Summenhäufigkeitskurve



Beispiel 4.7:

Bei einer Untersuchung der Lebensdauer (in Jahren) von 600 Fernsehgeräten eines bestimmten Typs ergaben sich die im folgenden aufgeführten Werte, deren Median (Zentralwert) zu bestimmen ist.

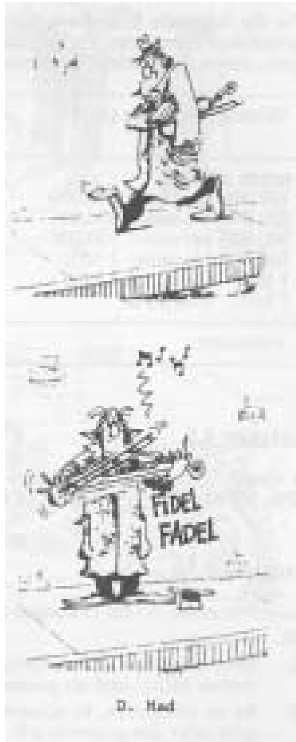
k	x_k *)	n_k	h_k	H_k
1	0 - 2	21	0,035	0,035
2	2 - 4	178	0,297	0,332
3	4 - 6	255	0,425	0,757
4	6 - 8	123	0,205	0,962
5	über 8	23	0,038	1,000

*) Lebensdauer von - bis unter - Jahre

Lösung 4.7:

Aus der Spalte H_k ist zu erkennen, dass k = 3 die Medianklasse ist, weil die bis zu Beginn der dritten Klasse erreichte kumulierte Häufigkeit 33,2% ist, während sie am Ende dieser Klasse 75,7% beträgt. Es gilt also mit der Interpolationsformel (Gl. 4.26):
Untergrenze der Medianklasse 4, Breite b_k = 2, relative Häufigkeit h_k = 0,425 und bei Beginn der Medianklasse H_{k-1} = 0,332. Also ist $\tilde{x}_{0,5}$ = 4 + 2(0,5 - 0,332)/0,425 = 4,791.

Beispiel 4.8: Der optimale Standort des Stehgeigers von Budapest



László Varga möchte den Bewohnern der Häuser A bis D der Bartók-Straße seine Sonate in d-moll op. 125 zu Gehör bringen. Dabei wünscht der Tonkünstler, dass alle 25 Familien möglichst gleich gut die Gelegenheit haben, die Sonate zu hören und zu würdigen.
Wo sollte sich Varga hinstellen, wenn die Straße wie folgt aussieht (Im i-ten Haus, das x_i Meter vom Beginn der Straße entfernt ist, wohnen n_i Familien):

x_i	n_i	N_i
0	3	3
10	4	7
20	1	8
30	2	10
35	3	13
50	2	15

Man erkläre anhand dieser Aufgabe auch die "Minimumeigenschaft" des Medians!

Lösung 4.8:

Um von allen potentiellen Hörern (und Zahlern) gleich weit entfernt zu sein, sollte sich Varga in den Medianpunkt stellen. Der Median hat die geringste Summe der absoluten Abweichungen. Er beträgt in diesem Fall $\tilde{x}_{0,5} = 20$, da bei n = 15 Familien der Median der 8-te [8=(15+1)/2] Wert ist. Das arithmetische Mittel beträgt 21,667. Mit d_{Zi} sollen die Abstände (absolute Abweichungen) vom Median bezeichnet werden, mit d_{ai} die Abstände vom Wert x = 22, mit d_{bi} die Abweichungen von x = 18 und mit d_{ci} die Abstände von x = 19:

		x = z = 20		x = 22		x = 18		x = 19	
x_i	n_i	d_{Zi}	$d_{Zi}n_i$	d_{ai}	$d_{ai}n_i$	d_{bi}	$d_{bi}n_i$	d_{ci}	$d_{ci}n_i$
0	3	20	60	22	66	18	54	19	57
10	4	10	40	12	48	8	32	9	36
20	1	0	0	2	2	2	2	1	1
30	2	10	20	8	16	12	24	11	22
35	3	15	45	13	39	17	51	16	48
50	2	30	60	28	56	32	64	31	62
$\sum$			225		227		227		226

Wie man sieht, ist die Summe der Abstände vom Median (20) kleiner als von den benachbarten Positionen (x = 18, 19 oder 22).

b) Quantile

Quantile teilen eine Verteilung in Abschnitte gleicher Häufigkeit ein, so beispielsweise:

- zwei **Terzile** T_1, T_2 in drei Abschnitte mit den relativen Häufigkeiten $h_1 = 1/3$ (für das Intervall $x < T_1$), $h_2 = 1/3$ (für das Intervall $T_1 \leq x < T_2$) und $h_3 = 1/3$ (für $x \leq T_2$)
- drei **Quartile** Q_1, $Q_2 = \tilde{x}_{0,5}$ (Median) und Q_3 in vier Abschnitte mit den relativen Häufigkeiten $h_1 = ... = h_4 = 1/4$
- vier **Quintile** in fünf Abschnitte mit $h_1 = ... = h_5 = 0{,}2$
- neun **Dezile** in zehn Abschnitte mit $h_1 = ... = h_{10} = 0{,}1$.

Das allgemeine Konzept ist das des Quantils (oder p-Quantils), wovon Terzile, Quartile, Quintile und Dezile Spezialfälle sind. Es handelt sich dabei um Lageparameter im allgemeinen Sinne. Nur das zweite Quartil, das identisch mit dem Median ist, kann als Mittelwert bezeichnet werden.

Def. 4.8: Quantil

Das Merkmal X sei mindestens ordinalskaliert. [c] bedeutet "ganze Zahl, die kleiner oder gleich c ist" (Gaußklammer).
Dann heißt die Maßzahl

$$(4.27)\quad \tilde{x}_p = \begin{cases} x_{([np+1])}, & \text{wenn np nicht ganzzahlig ist} \\ \tfrac{1}{2}(x_{[np]}+x_{[np+1]}), & \text{wenn np ganzzahlig ist} \end{cases}$$

p-Quantil ($0 < p < 1$).

Spezielle Quantile und Anzahl der Intervalle			
Zentile (Perzentile)	100	Quintile	5
Vigintile	20	Quartile	4
Dezile	10	Terzile	3

Bemerkungen zu Def.4.8:

1. Def. 4.8 ist eine Verallgemeinerung von Def. 4.7 für den Zentralwert, der das Quantil für $p = 0{,}5$ darstellt.

2. Die p-Quantile teilen die Daten in zwei Gruppen auf. Unterhalb von $\tilde{x}_p$ befinden sich höchstens $100 \cdot p$% der Merkmalswerte und oberhalb von $\tilde{x}_p$ befinden sich höchstens $100(1-p)$% der Merkmalswerte. Sofern np ganzzahlig ist, gelten diese Prozentsätze exakt. Gl. 4.27 kann ohne Schwierigkeiten auch bei gruppierten Daten angewendet werden.

3. Genau dann, wenn np nicht ganzzahlig ist, sind Quantile realisierte Merkmalswerte. Für ganzzahlige Werte np ist die Intervallmitte zwischen den beiden benachbarten Werten $x_{(np)}$ und $x_{(np+1)}$ der geordneten Reihe zu betrachten.

4. Klassierte Daten:
   Bei klassierten Daten können p-Quantile geometrisch leicht gedeutet werden. Ein p-Quantil ist derjenige Wert, bei dem die empirische Verteilungsfunktion H(x) den Wert p annimmt. Es kann im allgemeinen nur näherungsweise durch Interpolation bestimmt werden. Die Berechnungsformel ist analog zu Gl. 4.26. Ist k die Klasse, in die das p-Quantil $\tilde{x}_p$ fällt, dann ist $\tilde{x}_p$ näherungsweise durch

   (4.26a) $\tilde{x}_p = x'_{k-1} + b_k(p - H_{k-1})/h_k$

   gegeben.

5. Die p-Quantile sind Funktionswerte der inversen Verteilungsfunktion $G[H(x)] = G(H)$ (vgl. Kap.3, Def. 3.3). So erhält man beispielsweise den Median $Z = \tilde{x}_{0,5}$ durch die Lösung der Gleichung $Z = G(0{,}5)$ und die beiden Quartile Q_1 und Q_3 sind durch die Funktionswerte $Q_1 = G(0{,}25)$ und $Q_2 = G(0{,}75)$ der Funktion G(H) gegeben.

6. Aus Quantilen lassen sich durch Mittelung Lageparameter herleiten, etwa die

Quartilsmitte	$\frac{1}{2}(Q_1+Q_3)$

oder entsprechend die Dezilsmitte ½(D_1+D_9). Ein Grenzfall ist

$$\boxed{(4.28) \quad \bar{x}_m = ½(x_{min}+x_{max})}$$

bekannt als der "Mittelpunkt" [midpoint oder midrange] (des Streubereichs), der auch "midrange" genannt wird. Man kann $\bar{x}_m$ herleiten aus einer Minimierungsaufgabe: $\bar{x}_m$ ist das Minimum der maximalen Abstände der Werte x_v von einem festen Wert c, was in Aufgabe 4.10 demonstriert wird:

$$\max_v |x_v - \bar{x}_m| \le \max_v |x_v - c|.$$

Im Unterschied zur Quartilsmitte ist $\bar{x}_m$ sehr schnell durch extreme Werte verzerrt. Deshalb kann $\bar{x}_m$ keine Alternative zu weniger sensiblen Mittelwerten sein.

Beispiel 4.9:

Die Untersuchung eines Einzelhändlers über die Umsatzzahlen U_v (DM pro Tag) der 10 im Sortiment befindlichen Radiotypen ergab die folgenden Werte U_v: 350, 420, 780, 890, 975, 1010, 1170, 1230, 1680, und 1910

a) Man bestimme die Quartile Q_1 (das untere Quartil) und Q_3 (oberes Quartil) sowie die Quartilsmitte! Unterscheiden sich Median und Quartilsmitte?
b) Man verfahre entsprechend, wenn sich die Statistik nur auf (die ersten) acht Radiotypen erstreckt.

Lösung 4.9:

a) zehn Radiotypen
Unteres Quartil Q_1:
p = 0,25. Dann ist np = 2,5 (nicht ganzzahlig!), so dass nach Def. 4.8 x_p der Wert $x_{([np+1])}$ ist. Nun ist np + 1 = 3,5 . Die Gaußklammer bedeutet "ganze Zahl, die kleiner oder gleich 3,5 ist", also die Zahl 3. Folglich ist $Q_1 = x_{(3)} = 780$.
Oberes Quartil Q_3:
p = 0,75. Die entsprechende Betrachtung führt zu $Q_3 = x_{(8)} = 1230$. Die Quartilsmitte ist folglich 1005. Bei der Bestimmung des Zentralwerts (Medians) ist Def. 4.8 mit p = 0,5 anzuwenden: der Median ist somit $½\,[x_{(np)} + x_{(np+1)}] = ½\,[x_{(5)} + x_{(6)}] = ½\,(975 + 1010) = 992{,}5$ (im Unterschied zur Quartilsmitte, die 1005 beträgt).

b) acht Radiotypen

Q_1: In diesem Fall ist np = 8·0,25 = 2 also ganzzahlig, so dass Q_1 als Mittelwert aus $x_{(np)}$ und $x_{(np+1)}$, also aus $x_{(2)}$ = 420 und $x_{(3)}$ = 780 zu bestimmen ist (Q_1 ist also 600);
Q_3: das obere Quartil ist das ungewogene Mittel aus $x_{(6)}$ = 1010 und $x_{(7)}$ = 1170 also 1090. Die Quartilsmitte ist ½(600+1090) = 845 (der Median ist dagegen ½(890+975) = 932,5 also nicht identisch mit der Quartilsmitte).

Beispiel 4.10:

Man zeige für die folgenden Daten x: 10,12,15,18,22, dass für die midrange $\underset{v}{\max} x_m$ (Gl. 4.28) gilt: $\max |x_v - \underset{v}{\max} x_m| \leq \max |x_v - c|$ (mit c als einer von $\underset{v}{\max} x_m$ verschiedenen Konstanten).

Lösung 4.10:

In diesem Fall gilt $\bar{x}$ = ½(10+22) = 16. Die Abweichungen der einzelnen Werte von verschiedenen Werten für c betragen:

Daten	Abweichungen von c								
x_v	c=12	c=13	c=14	c=15	c=16	c=17	c=18	c=19	c=20
10	2	3	4	5	6	7	8	9	10
12	0	1	2	3	4	5	6	7	8
15	3	2	1	0	1	2	3	4	5
18	6	5	4	3	2	1	0	1	2
22	10	9	8	7	6	5	4	3	2
$\max \lvert x_v - c \rvert$	10	9	8	7	6	7	8	9	10

Wie man sieht ist die geringste Maximalabweichung 6. Das ist die Abweichung von c = $\bar{x}_m$ = 16.

c) Modus (dichtester Wert, häufigster Wert)

Def. 4.9: Modus

Existiert bei einer diskreten Variable (einem diskreten Merkmal) X mit den Merkmalsausprägungen x_i genau ein Merkmalswert $\bar{x}_M$, so dass

$$(4.29) \quad h(x = \bar{x}_M) = \max_i h(x_i),$$

so ist dieser Wert der Modus $\bar{x}_M$ [oder D] (oder der Modalwert, der dichteste- bzw. häufigste Wert).
Der Modus ist derjenige Merkmalswert, der in einer Häufigkeitsverteilung am häufigsten vorkommt.

Bemerkungen zu Def. 4.9:

1. Während zur Bestimmung des Medians aus Einzelwerten eine geordnete Reihe gebildet werden muss, sind zur Ermittlung des Modus die Einzelwerte zu gruppieren. Die Bestimmung des Modus ist auch sinnvoll, wenn eine Klassierung vorgenommen wurde (d.h. bei einer klassierten Verteilung). Die Lage (der Wert) des Modus (der Modalklasse) kann dann aber durch unterschiedliche Arten der Klassenbildung verschoben werden (vgl. auch Bem. Nr. 6).

2. Mit dem Modus verbindet man eine gewisse Vorstellung von Normalität und Üblichkeit. Mit "normal" ist meist der häufigste Wert (der Gipfel einer Verteilung) gemeint. Bei unklassierten Daten ist der Modus immer ein realisierter Wert, was bei den anderen Lageparametern nicht notwendig der Fall zu sein braucht, so dass er sich durch hohe Realitätsnähe auszeichnet.

4. Transformation und Skaleneigenschaft: Es gilt bei einer eineindeutigen Transformation $x^*_i = f(x_i)$ für den Modus der transformierten Werte $D^* = f(D)$. Der Modus ist also schon bei nur nominalskalierten Merkmalen ein sinnvolles Lagemaß.

4. Die Bestimmung des Modus macht wenig Sinn, wenn die Verteilung nicht eingipflig (unimodal) ist. Bei einer U-förmigen Häufigkeitsverteilung nennt man ein lokales Minimum auch Antimodus.

5. Ein extremer Fall der Abstandsmessung ist die Verabredung, den Wert $d_v = d(x_v,x)$ (d bedeutet Distanz zwischen x_v und x) wie folgt zu definieren:
   $d_v = 0$ wenn $x_v = x$ und $d_v = 1$ wenn $x_v \neq x$
   (x ist ein beliebiger Wert für das Merkmal X [eine beliebige Merkmalsausprägung])
   Der Modus ist dann derjenige Wert von x, für den gilt $\Sigma d_v = \min\limits_x$, denn ist $x = \bar{x}_M$, so wird $d_v = 0$ häufiger als bei allen anderen Werten für x vergeben.
   Das zeigt erneut, dass man Mittelwerte aufgrund von Extremwerten von "Gesamtabständen" definieren kann: das arithmetische Mittel ist wegen Satz 4.2 derjenige Wert x, für den die Größe $\Sigma(x_v - x)^2$, also die Summe der quadrierten Abstände minimal ist.

6. Klassierte Daten
   Bei einer klassierten Verteilung begnügt man sich im allgemeinen mit einer Angabe der Modalklasse d, also der Klasse, deren (relative) Häufigkeit die größte der Häufigkeitsverteilung ist. Als Modus gilt dann die Klassenmitte m_d der Modalklasse. Seltener bestimmt man den Modus durch Interpolation. Sofern die modale Klasse d sowie die beiden angrenzenden Klassen die gleiche Breite b_d haben, ist der Modus:

$$\boxed{(4.30) \qquad \bar{x}_M = D = x'_d + (A \cdot b_d)/(A+B)}$$

mit der Klassenuntergrenze x'_d der Modalklasse und den beiden Differenzen von relativen Häufigkeiten der benachbarten Klassen $A = h_d - h_{d-1}$ und $B = h_d - h_{d+1}$. Der Modus D rückt damit näher an die jeweils stärker besetzte der beiden benachbarten Klassen (in Abb.4.10 an die Klasse d+1).

Beispiel 4.11:

Man bestimme den Modus nach Gl. 4.30 für das Beispiel 4.2!

Lösung 4.11:

Die Modalklasse ist die Klasse 5. Die Klassenbreiten der 5-ten und angrenzenden Klassen sind b_d = 400DM. Die Untergrenze der Modalklasse beträgt 2200DM. Ferner ist A = (37-31)/150 und B = 15/150, so dass man $\bar{x}_M$ = D = 2200 + (6/21)400 = 2314,29 erhält.

Abb.4.10: Interpolation des Modus