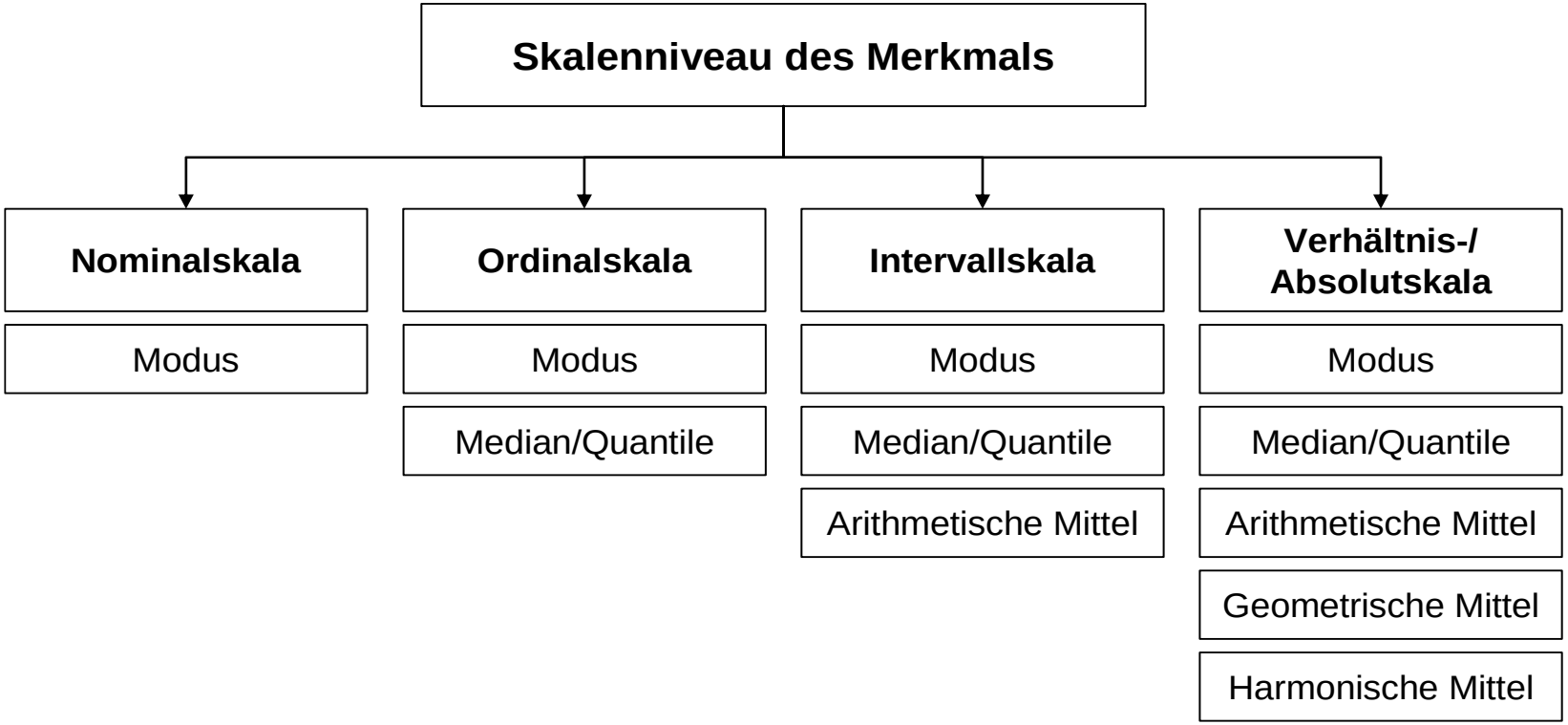
4. Lagemaße

- Häufigkeitsverteilungen: Detaillierter Einblick in die Datenstruktur
- Maßzahlen (Kennzahlen): Beschreibung bestimmter Eigenschaften von Daten durch eine Zahl

Eine **Maßzahl** ist eine Kennzahl zur Beschreibung der Merkmalswerte.

Lagemaße geben über die zentrale Tendenz eines Merkmals Auskunft.



Kriterien für die Auswahl von Lagemaßen:

1. **Skalenniveau**:
   Nicht alle Lagemaße sind anwendbar, wenn das Merkmal nicht die Anforderungen an das Skalenniveau erfüllt.

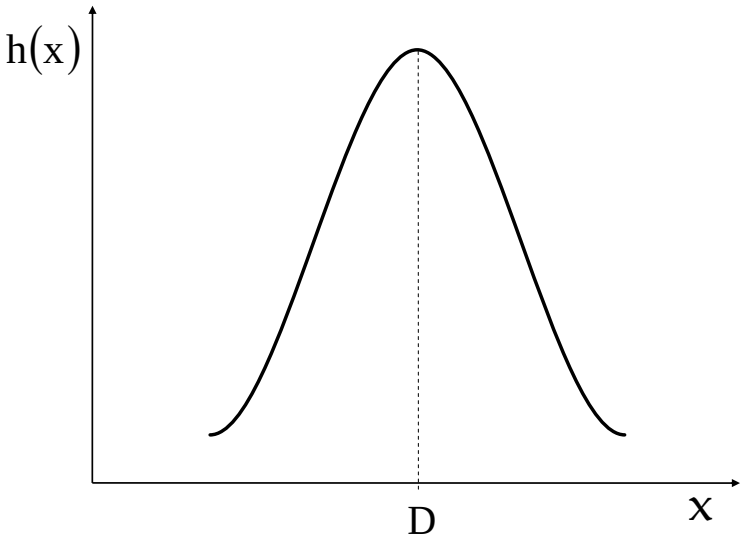
2. **Aussagezweck**:
   Interessiert ein Mittelwert oder beispielsweise das erste Quartil, das die kleinsten 25% der Merkmalswerte von den größten 75% der Daten trennt?

3. **Sachlogik**:
   Ist z. B. das arithmetische oder geometrische Mittel zur Bestimmung einer durchschnittlichen Wachstumsrate der sachlich korrekte Mittelwert?

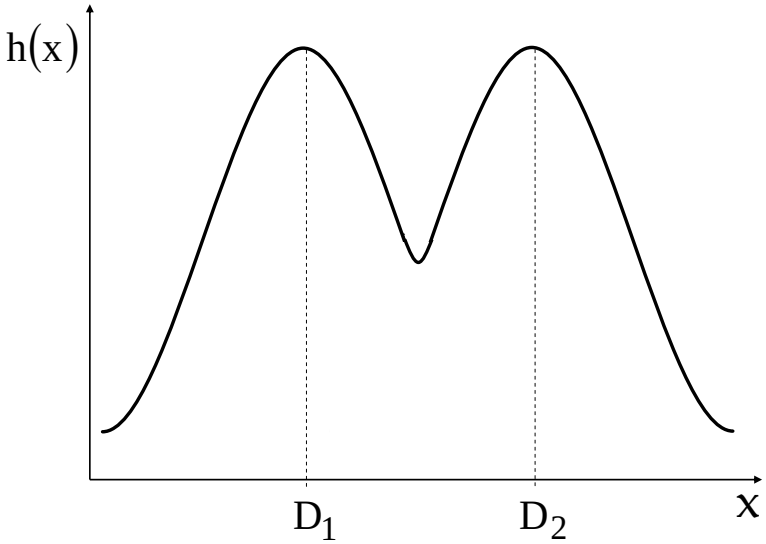
4.1 Modus

Der **Modus** (**Modalwert**) D ist als häufigster oder typischer Wert zu interpretieren.

- Anwendbar bei jedem beliebigem Skalenniveau
- Fällt mit dem Gipfel der Häufigkeitsverteilung zusammen
    - Bei zwei Maxima: Bimodale Verteilung (Modalwerte D_1 und D_2)
    - Bei mehreren Maxima: Multimodale Verteilung (Modalwerte D_1, D_2, …)



a) unimodal



b) bimodal

Berechnung bei Einzelwerten und unklassierter Häufigkeitsverteilung

Der Modus D ist die Merkmalsausprägung mit der maximalen (absoluten und relativen) Häufigkeit.

- D = x_l^*
- l als Laufindex, der mit der maximalen Häufigkeit korrespondiert
- Maximale absolute Häufigkeit: $n_l = \max\{n_j\}$
- Maximale relative Häufigkeit: $h_l = \max\{h_j\}$

Beispiel: Lieblingsmusik

Fünf Studierende werden nach ihrer Lieblingsmusik gefragt. Dabei liegen folgende Einzelwerte vor:

x_1 = Volksmusik, x_2 = Volksmusik, x_3 = Jazz, x_4 = Volksmusik, x_5 = Klassik

Die größte absolute Häufigkeit weist die Volksmusik auf, da diese Ausprägung dreimal gemessen wird (übrige Musikrichtungen jeweils einmal):

Absolute Häufigkeit			Berechnung des Modus
j	x_j^*	n_j	I = 1, da $n_I = n_1 = \max\{n_j\}$
1	Volksmusik	3	D = $x_I^* = x_1^*$ = Volksmusik
2	Jazz	1	
3	Klassik	1	
∑		n = 5	

Interpretation: Die meisten befragten Studierenden hören am liebsten Volksmusik.

Beispiel (Fortsetzung): Kinderzahl

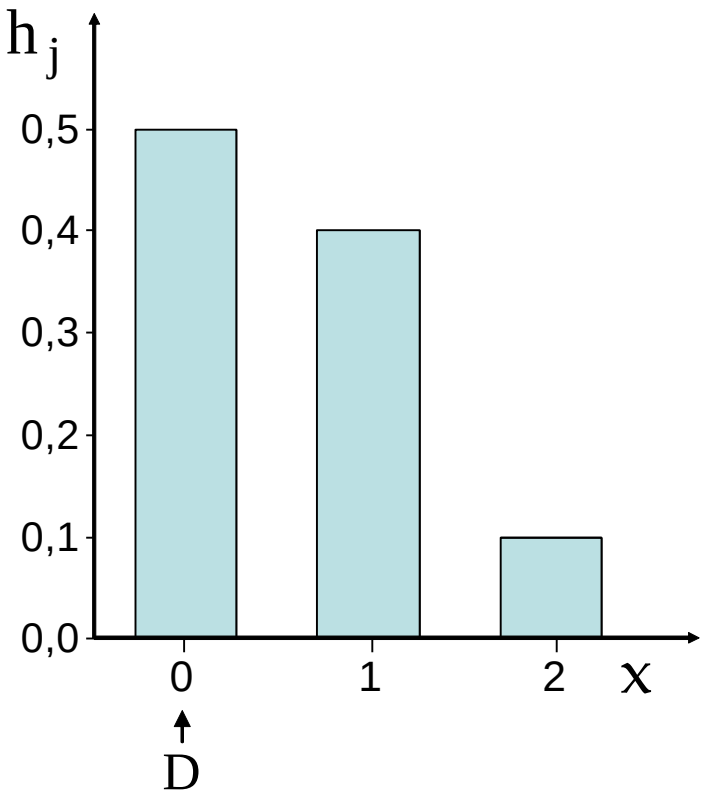
Beim Merkmal „Kinderzahl“ ist bekannt, dass fünf Studierende kein Kind, vier Studierende ein Kind und eine Befragte zwei Kinder besitzen. Wie groß ist der Modalwert?

Berechnung des Modus	Grafische Darstellung

j	x_j^*	n_j	h_j
1	0	5	0,5
2	1	4	0,4
3	2	1	0,1
$\sum$		$n = 10$	1

$l = 1$, da $h_l = h_1 = \max\{h_j\}$

$D = x_l^* = x_1^* = 0$



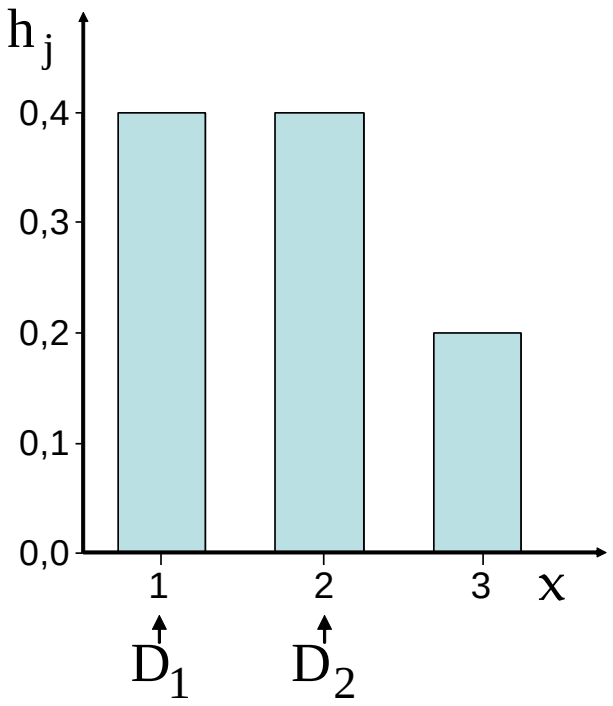
Beispiel: Haushaltsgröße

Die Häufigkeitsverteilung vom Merkmal „Haushaltsgröße“ weist zwei Gipfel auf (bimodale Verteilung). Somit müssen zwei Modalwerte ausgewiesen werden:

Berechnung des Modus	Grafische Darstellung

j	x_j^*	n_j	h_j
1	1	2	0,4
2	2	2	0,4
3	3	1	0,2
$\sum$		n = 5	1

$D_1 = x_1^* = 1$ und $D_2 = x_2^* = 2$, da die ersten beiden Merkmalsausprägungen die maximale Häufigkeit aufweisen.



Berechnung des Modus bei klassierter Häufigkeitsverteilung

Bei klassierten Häufigkeitsverteilungen liegt der Modus D in der Klasse mit der maximalen Häufigkeitsdichte (= Modalklasse k*).

- Klassenmitte der k-ten Klasse:

$$m_k = \frac{1}{2}(x'_{k-1} + x'_k)$$

- D wird durch die **Klassenmitte der Modelklasse** repräsentiert:

$$D = m_{k^*} \quad \text{mit} \quad m_{k^*} = \frac{1}{2}(x'_{k^*-1} + x'_{k^*})$$

- k*: Laufindex, der mit der maximalen Häufigkeitsdichte korrespondiert:

$$d_{k^*} = max\{d_k\}$$

Beispiel (Fortsetzung): Rendite

Den Modalwert der Rendite erhält man als Klassenmitte der Modalklasse:
k*=3, weil die dritte Klasse die maximale Häufigkeitsdichte aufweist.

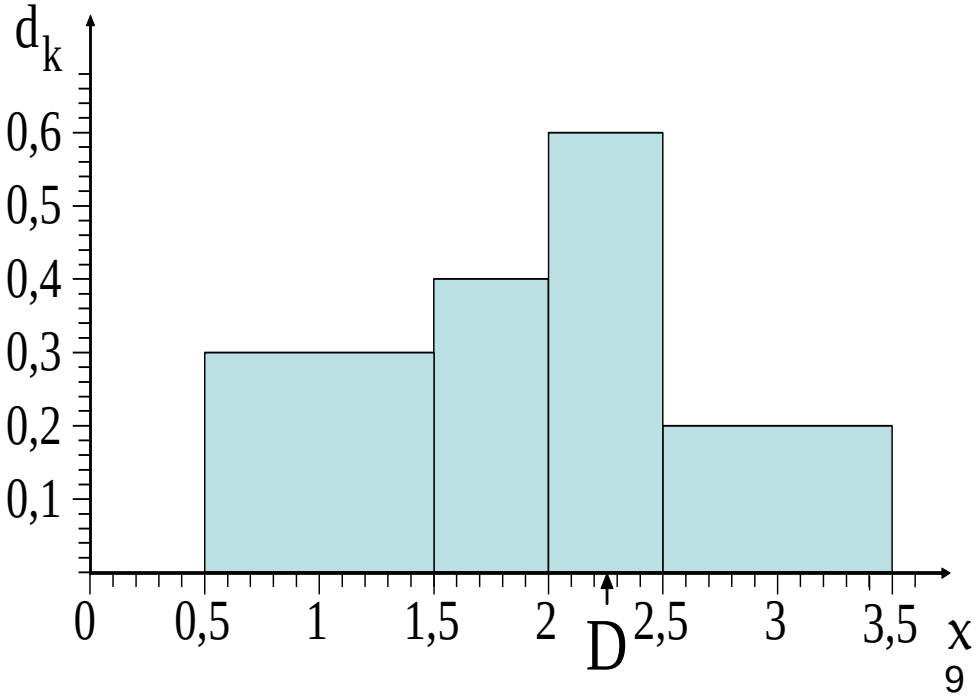
$$D = \frac{1}{2} \cdot \left(x'_{k^*-1} + x'_{k^*} \right) = \frac{1}{2} \cdot \left(x'_2 + x'_3 \right) = \frac{1}{2} \cdot (2 + 2{,}5) = 2{,}25$$

Interpretation: Die meisten Aktien verzeichnen eine (näherungsweise) Rendite von 2,25%. Hierbei handelt es sich um eine typische Aktienrendite.

Häufigkeitsdichten

k	von über... bis zu…	h_k	d_k
1	0,5 – 1,5	0,3	0,3
2	1,5 – 2	0,2	0,4
3	2 – 2,5	0,3	0,6
4	2,5 – 3,5	0,2	0,2
Σ		1	

Grafische Darstellung



Beispiel: Wein

Auf einem Weingut wird der Alkoholanteil von verschiedenen Weinen gemessen. 40 Weine haben einen Alkoholanteil von über 8% bis zu 10%, 20 Weine von über 10% bis zu 13% und 40 Weine von über 13% bis zu 15 %. Was ist der typische Alkoholanteil der Weine?

Zu berechnen ist der Modus. Die Häufigkeitsdichten werden in einer Häufigkeits-tabelle bestimmt:

k	von über… bis zu …	n_k	h_k	b_k	d_k
1	8 – 10	40	0,4	2	0,200
2	10 – 13	20	0,2	3	0,067
3	13 – 15	40	0,4	2	0,200
∑		n = 100	1		

Da zwei Klassen ($k_1^* = 1$ und $k_3^* = 3$) mit maximaler Häufigkeitsdichte vorhanden sind, liegen zwei Modalwerte (typische Werte) vor:

$$D_1 = \frac{1}{2} \cdot \left(x_0' + x_1' \right) = \frac{1}{2} \cdot (8 + 10) = 9 \qquad \text{(1. Modalwert)}$$

$$D_2 = \frac{1}{2} \cdot \left(x_2' + x_3' \right) = \frac{1}{2} \cdot (13 + 15) = 14 \qquad \text{(2. Modalwert)}$$

Eigenschaften des Modus:

- Anschaulicher Mittelwert (häufig beobachteter (typischer) Wert)
- Kann für jedes Skalenniveau berechnet werden
- Nutzt bei ordinalskalierten und metrischen Merkmalen nur einen geringen Teil der Informationen aus
- Daher vor allem für nominalskalierte Merkmale zu verwenden
- Bei flachen Häufigkeitsverteilungen wenig aussagekräftig

4.2 Median

- Der **Median** (Zentralwert) $\tilde{x}$ teilt die Merkmalswerte in zwei gleich große Hälften:
    - Mindestens 50 % der Merkmalswerte sind kleiner oder gleich dem Median
    - Mindestens 50 % der Merkmalswerte sind größer oder gleich dem Median
- Der Median lässt sich nur dann sinnvoll berechnen, wenn die Merkmalsausprägungen in aufsteigender Reihenfolge geordnet werden können

→ Voraussetzung: Mindestens ordinalskalierte Merkmale

Berechnung des Medians bei Einzelwerten

1. Schritt: Erstellung einer geordneten Reihe (in aufsteigender Reihenfolge):

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)} \quad \text{mit} \quad x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

2. Schritt: Bestimmung des Medians aus der geordneten Reihe

- **Falls n ungerade**:
  - Bei einem ungeraden n gibt es genau einen mittleren Wert, der die Ordnungszahl (n+1)/2 besitzt.
  - $\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$

- **Falls n gerade**:
  - Bei einem geraden n sind zwei mittlere Werte mit den Ordnungszahlen n/2 und n/2+1 vorhanden. Alle Werte zwischen den beiden mittleren Werten erfüllen die Medianeigenschaft. Als Median verwendet man die Intervallmitte.
  - $\tilde{x} = \frac{1}{2} \cdot \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$

Beispiel: Krankheitstage

Ein Unternehmen misst die Krankheitstage von fünf Mitarbeiter:innen:

$x_1 = 2$, $x_2 = 3$, $x_3 = 8$, $x_4 = 5$, $x_5 = 1$

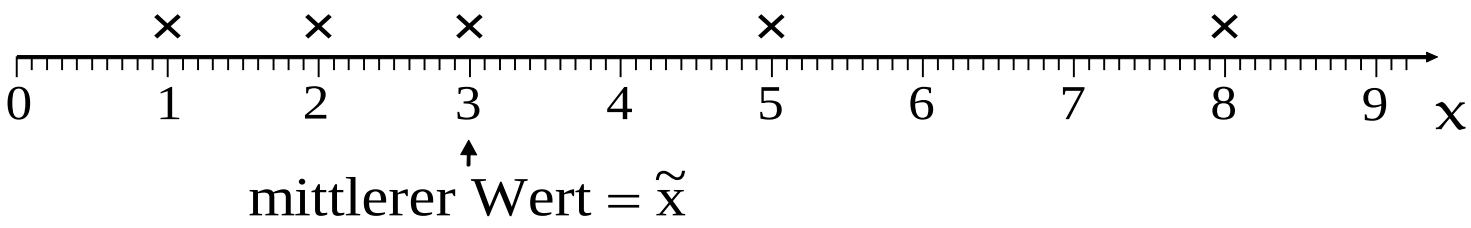
Zu berechnen ist der Median:

1. Geordnete Reihe:

   $x_{(1)} = 1$, $x_{(2)} = 2$, $x_{(3)} = 3$, $x_{(4)} = 5$, $x_{(5)} = 8$

2. n = 5 (ungerade):

   $$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{5+1}{2}\right)} = x_{\left(\frac{6}{2}\right)} = x_{(3)} = 3$$



Wie ändert sich der Median, wenn ein sechster Mitarbeiter mit sieben Krankheitstagen ebenfalls einbezogen wird?

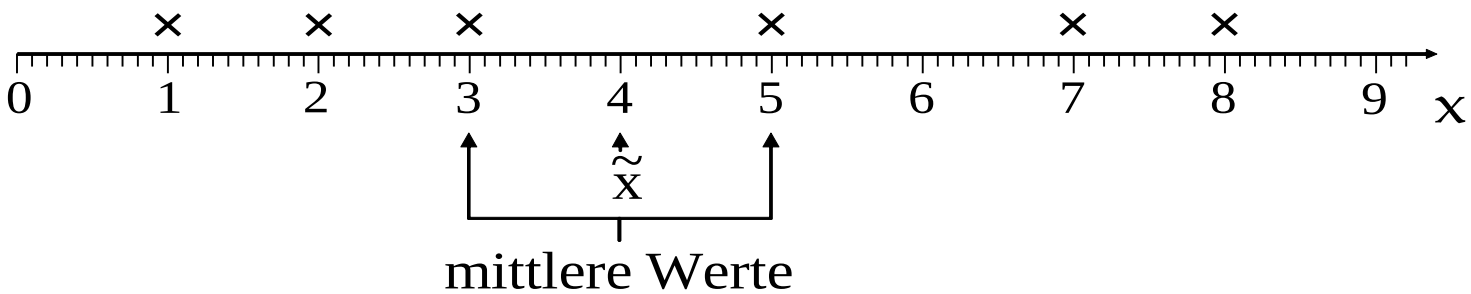
Dann gibt es zwei mittlere Werte, deren Durchschnitt zu bilden ist:

1. Geordnete Reihe:

   $x_{(1)} = 1$, $x_{(2)} = 2$, $x_{(3)} = 3$, $x_{(4)} = 5$, $x_{(5)} = 7$, $x_{(6)} = 8$

2. n = 6 (gerade):

$$\tilde{x} = \frac{1}{2} \cdot \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) = \frac{1}{2} \cdot \left(x_{\left(\frac{6}{2}\right)} + x_{\left(\frac{6}{2}+1\right)}\right) = \frac{1}{2} \cdot \left(x_{(3)} + x_{(4)}\right) = \frac{1}{2} \cdot (3+5) = 4$$



Beispiel: Pharmakonzerne

Von 10 ausgewählten Pharmakonzernen liegt der Umsatz vor (in Mrd. $):

17,3; 16,6; 14,7; 27,0; 17,2; 21,6; 13,6; 40,3; 12,4; 11,7

1. Geordnete Reihe:

   $x_{(1)} = 11{,}7$; $x_{(2)} = 12{,}4$; $x_{(3)} = 13{,}6$; $x_{(4)} = 14{,}7$; $x_{(5)} = 16{,}6$;

   $x_{(6)} = 17{,}2$; $x_{(7)} = 17{,}3$; $x_{(8)} = 21{,}6$; $x_{(9)} = 27{,}0$; $x_{(10)} = 40{,}3$

2. n = 10 (gerade):

$$\tilde{x} = \frac{1}{2}\left(x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}\right) = \frac{1}{2}\left(x_{(5)} + x_{(6)}\right) = \frac{1}{2}(16{,}6 + 17{,}2) = 16{,}9$$

Interpretation: Der Median der Umsätze beträgt 16,9 Mrd. $. 50% der Pharmakonzerne verzeichnen einen geringeren Umsatz als 16,9 Mrd. $. Ebenso weisen 50% der Konzerne einen höheren Umsatz als 16,9 Mrd. $ auf.

Berechnung des Medians bei unklassierter Häufigkeitsverteilung

1. Bestimmung der kumulierten relativen Häufigkeiten (zumindest bis zur 50%-Marke): $H_1, H_2, \ldots, H_m = 1$
2. Bestimmung des Medians bei Orientierung an der 50%-Marke der kumulierten relativen Häufigkeiten
   - Falls die 50%-Marke bei einer Merkmalsausprägung durchlaufen wird, stimmt der Median mit dieser Merkmalsausprägung überein.
   - Falls die 50%-Marke exakt bei einer Merkmalsausprägung erreicht wird, ist diese mit der nachfolgenden Merkmalsausprägung zu mitteln.

Berechnungsformel:

$$\tilde{x} = \begin{cases} x_l^* \text{ falls } H_l > 0{,}5 \text{ und } H_{l-1} < 0{,}5 \\ \frac{1}{2}(x_l^* + x_{l+1}^*) \text{ falls } H_l = 0{,}5 \end{cases}$$

wobei der Laufindex l der Index der Merkmalsausprägung ist, bei der die relativen kumulierten Häufigkeiten 0,5 erreichen oder durchlaufen.

Beispiel: Callcenter

In einem Callcenter werden die eingehenden Anrufe pro Minute elektronisch registriert. Für vier Untersuchungsperioden sind folgende Werte gemessen worden:

$x_1 = 10$, $x_2 = 11$, $x_3 = 13$, $x_4 = 10$

Der Median soll aus der unklassierten Häufigkeitsverteilung bestimmt werden:

j	x_j^*	n_j	h_j	H_j
1	10	2	2/4 = 0,5	0,5
2	11	1	1/4 = 0,25	0,5 + 0,25 = 0,75
3	13	1	1/4 = 0,25	0,75 + 0,25 = 1,0
∑		n = 4	1	

- Der Laufindex l ist gleich 1, weil die relativen kumulierten Häufigkeiten bei der ersten Merkmalsausprägung 0,5 erreichen.
- Weil $H_l = H_1 = 0{,}5$ ist, müssen die Merkmalsausprägungen x_1^* und x_2^* gemittelt werden:

$$\tilde{x} = \frac{1}{2}(x_1^* + x_2^*) = \frac{1}{2}(10 + 11) = 10{,}5$$

Beispiel: Statistik-Klausur

Bei einer Statistik-Klausur im Schwerpunktstudium liegen folgende Benotungen vor:

$x_1 = 5$, $x_2 = 3$, $x_3 = 1$, $x_4 = 5$, $x_5 = 3$

Der Median soll aus der unklassierten Häufigkeitsverteilung bestimmt werden:

j	x_j^*	n_j	h_j	H_j
1	1	1	1/5 = 0,2	0,2
2	3	2	2/5 = 0,4	0,2 + 0,4 = 0,6
3	5	2	2/5 = 0,4	0,6 + 0,4 = 1,0
∑		n = 5	1	

Hier gilt:

- Der Laufindex l ist gleich 2, weil die relativen kumulierten Häufigkeiten bei der zweiten Merkmalsausprägung 0,5 durchlaufen (überschreiten).

- Da $H_l = H_2 = 0{,}6 > 0{,}5$ und $H_{l-1} = H_1 = 0{,}2 < 0{,}5$ sind, muss keine Mittelung durchgeführt werden:

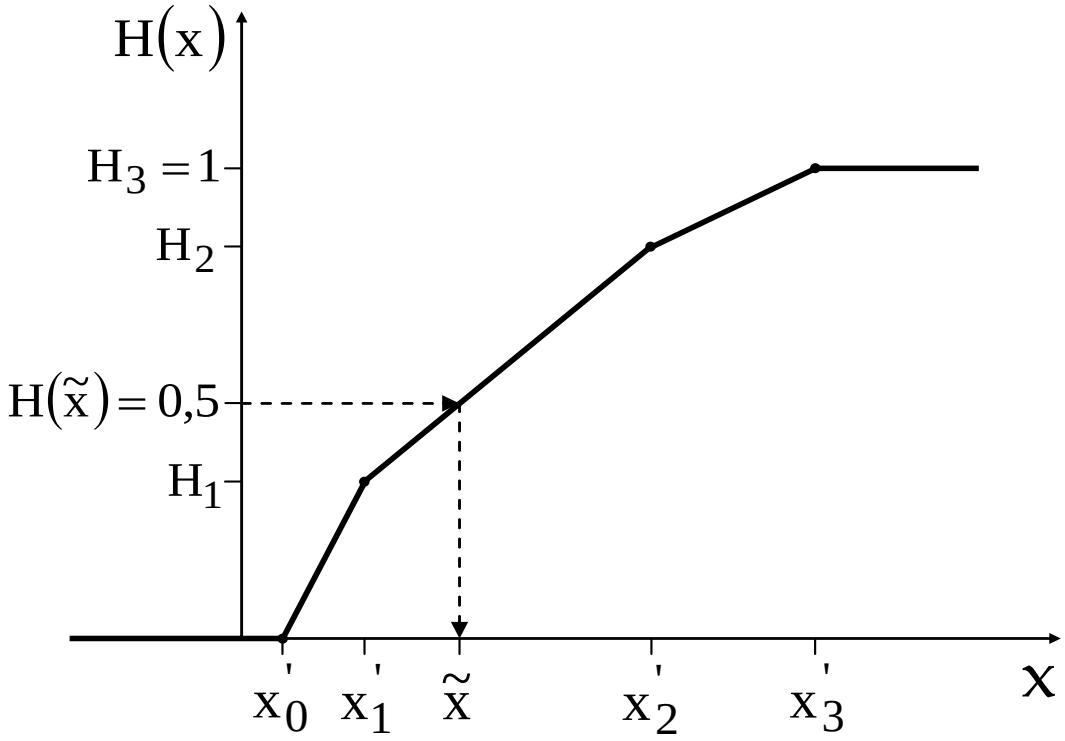
$$\tilde{x} = x_2^* = 3$$

Berechnung des Medians bei klassierter Häufigkeitsverteilung

- Wird in einer Klasse die 50%-Trennmarke der relativen kumulierten Klassenhäufigkeiten durchlaufen, dann liegt der Median in dieser Klasse (k*: Medianklasse).

- Welcher Wert in der Medianklasse soll als Median angegeben werden?

- Geht man von einer Gleichverteilung der Merkmalswerte innerhalb der Medianklasse aus, lässt sich der Median durch Interpolation bestimmen:



1. Schritt: Bestimmung der Medianklasse

- Medianklasse k*: Klasse, in der die 50%-Marke der kumulierten relativen Häufigkeiten durchlaufen wird
- In vorheriger Abbildung: $k^* = 2$ (Medianklasse = zweite Klasse)

2. Schritt: Approximative Berechnung des Medians

- Annahme: Gleichverteilung der Merkmalswerte innerhalb der Medianklasse
- Berechnungsformel (Herleitung siehe Lehrbuch):

$$\tilde{x} = x'_{k^*-1} + \frac{0{,}5 - H_{k^*-1}}{h_{k^*}} \cdot b_{k^*}$$

  wobei k* die Klasse ist, bei der die relativen kumulierten Klassenhäufigkeiten 0,5 durchlaufen (überschreiten).
- In vorheriger Abbildung:

$$\tilde{x} = x'_1 + \frac{0{,}5 - H_1}{h_2} \cdot b_2$$

Beispiel: IT-Unternehmen

Ein IT-Unternehmen misst die Anzahl der Zugriffe auf die Website pro Tag. Bei 20% der Messungen werden von über 1 bis zu 2 Mio. Zugriffe, bei 40% der Messungen von über 2 Mio. bis zu 3 Mio. Zugriffe und bei den restlichen Messungen von über 3 Mio. bis zu 3,5 Mio. Zugriffe registriert.

Es ist k*=2, da die kumulierten relativen Häufigkeiten H_k die 50%-Marke in der zweiten Klasse durchlaufen (überschreiten).

$$\tilde{x} = x'_{k^*-1} + \frac{0{,}5 - H_{k^*-1}}{h_{k^*}} \cdot b_{k^*} = x'_1 + \frac{0{,}5 - H_1}{h_2} \cdot b_2 = 2 + \frac{0{,}5 - 0{,}2}{0{,}4} \cdot 1 = 2{,}75$$

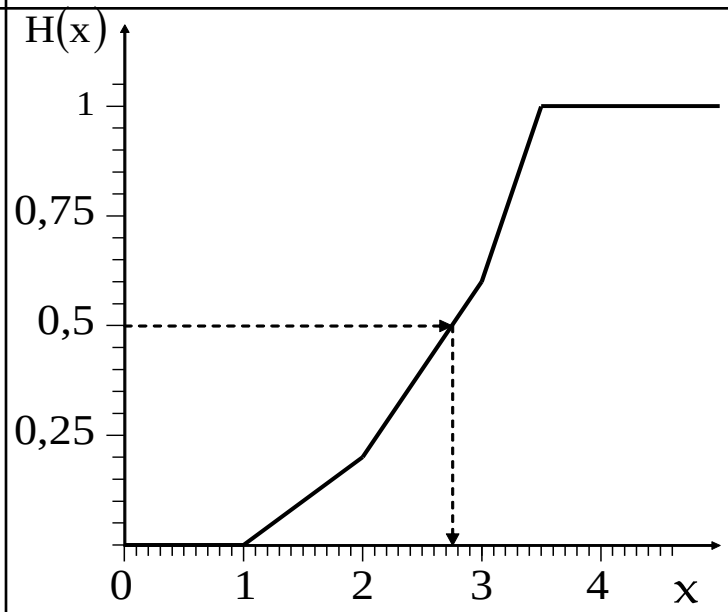
Kumulierte relative Häufigkeiten

k	von über bis zu…	b_k	h_k	H_k
1	1 – 2	1	0,2	0,2
2	2 – 3	1	0,4	0,6
3	3 – 3,5	0,5	0,4	1
∑			1	

$$H_2 = H_1 + h_2 = 0{,}2 + 0{,}4 = 0{,}6$$

$$b_1 = x'_1 - x'_0 = 2 - 1 = 1$$

Grafische Darstellung



Eigenschaften des Medians

1. Unempfindlichkeit (Resistenz) gegenüber Ausreißern

Beispiel:

Ein Souvenirladen in Dresden hat in den Jahren vor dem Elbhochwasser relativ konstante Umsatzzahlen im September ausgewiesen: 3200€, 3000€, 3250€, 3100€. Im Jahr des Elbhochwassers sank der Umsatz im angegebenen Monat auf 900€.

Median mit Ausreißer:

- $x_{(1)} = 900$; $x_{(2)} = 3000$; $x_{(3)} = 3100$; $x_{(4)} = 3200$; $x_{(5)} = 3250$
- $n = 5 \rightarrow \tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{6}{2}\right)} = x_{(3)} = 3100$

Median ohne Ausreißer:

- $x_{(1)} = 3000$; $x_{(2)} = 3100$; $x_{(3)} = 3200$; $x_{(4)} = 3250$
- $n = 4 \rightarrow \tilde{x} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) = \frac{1}{2}\left(x_{(2)} + x_{(3)}\right) = 3150$

Der Median verändert sich nur um 50€, wenn der Ausreißerwert berücksichtigt wird.

2. Minimumeigenschaft des Medians:

Die Summe der absoluten Abweichungen um eine beliebige reelle Zahl c

$$\sum_{i=1}^{n} |x_i - c|$$

nimmt ein Minimum an, wenn $c = \tilde{x}$ ist.

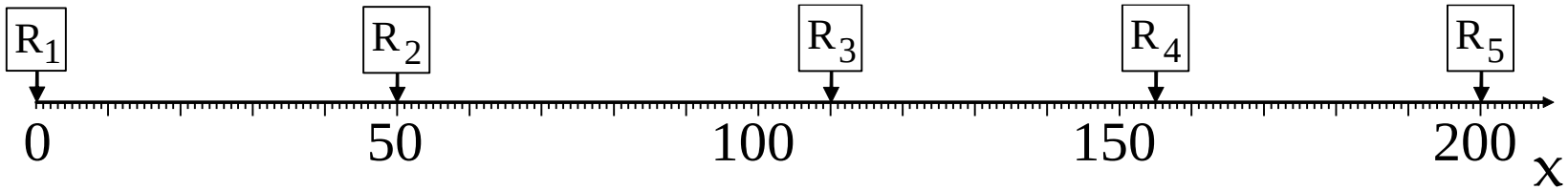
Es ist also

$$\sum_{i=1}^{n} |x_i - \tilde{x}| \leq \sum_{i=1}^{n} |x_i - c|$$

- Die Minimumeigenschaft ist beispielsweise dann relevant, wenn ein Ort bestimmt werden soll, von dem die Entfernung zu anderen Orten minimal ist (Zentralortproblem).
- Bedingungen:
  - Alle Orte liegen auf einer Strecke
  - Nur die betrachtete Strecke kommt in Frage, d.h. keine Alternativrouten sind möglich

Beispiel: Zentralortproblem

Ein Reinigungsunternehmen hat sich auf Autobahnraststätten spezialisiert. Der Unternehmenschef möchte einen neuen Standort für das Unternehmen finden, der minimale Anfahrtswege zu den Raststätten aufweist. Jede Reinigungsgruppe trifft sich am Sitz des Unternehmens, fährt jeweils eine Raststätte an und beendet die Tätigkeit am Sitz der Unternehmens. Jede Raststätte wird gleich häufig (einmal pro Tag) gereinigt. Die Raststätten R_1 und R_5 bilden jeweils die beiden äußeren Punkte. Die Entfernung zu R_1 beträgt bei R_2 50 km, bei R_3 110 km, bei R_4 155 km und bei R_5 200 km. Wo sollte der Sitz der Firma errichtet werden, um die Fahrwege zu minimieren?



- $x_{(1)} = 0$; $x_{(2)} = 50$; $x_{(3)} = 110$; $x_{(4)} = 155$; $x_{(5)} = 200$
- $n = 5 \rightarrow \tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{6}{2}\right)} = x_{(3)} = 110$
- Bei R_3, also 110 km entfernt von R_1, muss der Sitz des Unternehmens errichtet werden, um die Fahrtwege zu minimieren.

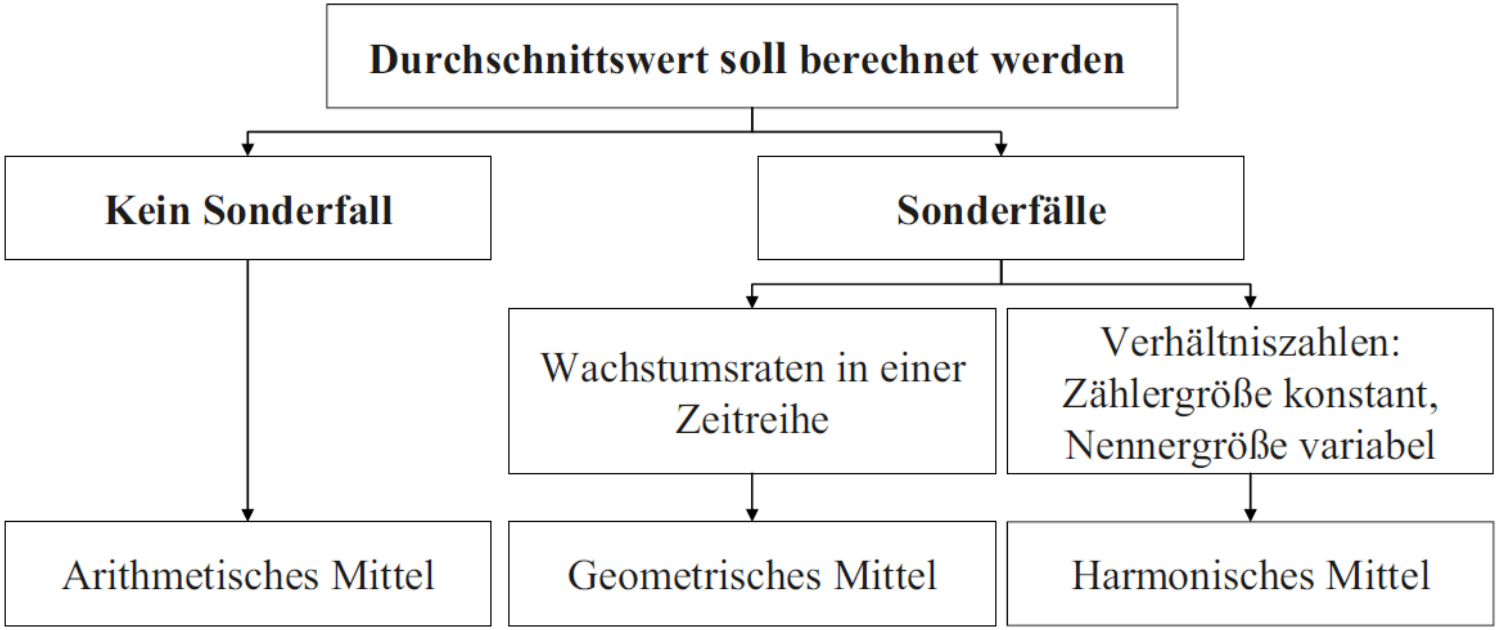
4.3 Arithmetisches Mittel

Allgemein ist das arithmetische Mittel ein Durchschnittswert, bei dem die Merkmalssumme auf die Anzahl der Merkmalsträger bezogen wird.

- Populärer „Durchschnitt“
- Bei der Berechnung werden Einzelwerte oder Merkmalsausprägungen aufsummiert

→ Voraussetzung: Metrisch skalierte Merkmale

Anwendungen verschiedener Durchschnittswerte:



Berechnung des arithmetischen Mittels bei Einzelwerten

- Quotient aus Merkmalssumme und der Anzahl statistischer Einheiten:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

- Gewicht jedes Einzelwerts: 1/n → ungewogenes arithmetisches Mittel

Beispiel (Fortsetzung): Callcenter

Unter Verwendung der Einzelwerte für die eingehenden Anrufe im Callcenter der vier Perioden, $x_1 = 10$, $x_2 = 11$, $x_3 = 13$, $x_4 = 10$, erhält man das folgende arithmetische Mittel:

$$\bar{x} = \frac{1}{4}\sum_{i=1}^{4} x_i = \frac{1}{4}(10 + 11 + 13 + 10) = 11$$

Berechnung des arithmetischen Mittels bei unklassierter Häufigkeitsverteilung

Summierung der mit den Häufigkeiten gewichteten Merkmalsausprägungen

$$\bar{x} = \frac{1}{n}(x_1^* + \cdots + x_1^* + x_2^* + \cdots + x_2^* + \cdots + x_m^* + \cdots + x_m^*)$$

- Mit absoluten Häufigkeiten:

$$\bar{x} = \frac{1}{n}(x_1^* \cdot n_1 + x_2^* \cdot n_2 + \cdots + x_m^* \cdot n_m) = \frac{1}{n} \cdot \sum_{j=1}^{m} x_j^* \cdot n_j$$

- Mit relativen Häufigkeiten:

$$\bar{x} = (x_1^* \cdot h_1 + x_2^* \cdot h_2 + \cdots + x_m^* \cdot h_m) = \sum_{j=1}^{m} x_j^* \cdot h_j$$

→ Gewogenes arithmetisches Mittel

Beispiel (Fortsetzung): Callcenter

j	x_j^*	n_j	h_j
1	10	2	0,5
2	11	1	0,25
3	13	1	0,25
∑		n = 4	1

- Berechnung mit absoluten Häufigkeiten:

$$\bar{x} = \frac{1}{4} \cdot \sum_{j=1}^{3} x_j^* \cdot n_j = \frac{1}{4}(10 \cdot 2 + 11 \cdot 1 + 13 \cdot 1) = \frac{1}{4}(20 + 11 + 13) = 11$$

- Berechnung mit relativen Häufigkeiten:

$$\bar{x} = \sum_{j=1}^{3} x_j^* \cdot h_j = (10 \cdot 0{,}5 + 11 \cdot 0{,}25 + 13 \cdot 0{,}25) = 11$$

→ Gleiches Ergebnis wie beim ungewogenen arithmetischen Mittel

Beispiel: Beratungsleitungen

Ein Unternehmen möchte wissen, wie häufig Beratungsleistungen von den Kunden durchschnittlich in Anspruch genommen werden. Hierfür hat es folgende Daten erhoben:

Inanspruchnahme von Beratungsleistungen	0	1	2	3
Anteil der Kunden	0,15	0,30	0,35	0,20

Zu berechnen ist das arithmetische Mittel:

$$\bar{x} = \sum_{j=1}^{4} x_j^* \cdot h_j = (0 \cdot 0{,}15 + 1 \cdot 0{,}30 + 2 \cdot 0{,}35 + 3 \cdot 0{,}20) = 1{,}60$$

→ Im Durchschnitt lassen sich die Kunden also 1,6-mal beraten.

Berechnung des arithmetischen Mittels bei klassierter Häufigkeitsverteilung (I)

Nutzung der Aggregationseigenschaft: Berechnung des gesamten arithmetischen Mittels durch Aggregation (Zusammenfassung) der arithmetischen Mittel von Teilgesamtheiten.

1. Fall: Vollständige Information

- Berechnung des arithmetischen Gesamtmittels $\bar{x}$ unter Verwendung der arithmetischen Klassenmittel $\overline{x_k}$:

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{p} \bar{x}_k \, n_k = \sum_{k=1}^{p} \bar{x}_k \, h_k$$

- Voraussetzung für Anwendung: Die arithmetischen Klassenmittel $\overline{x_k}$ …
  - … sind gegeben
  - … lassen sich aus den Einzelwerten der k-ten Klasse berechnen:

$$\bar{x}_k = \frac{\sum_{i=1}^{n_k} x_{ik}}{n_k} = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ik}$$

(x_{ik} sind die Einzelwerte der k-ten Klasse)

Beispiel: Marktforschung

Eine Marktforscherin erhebt des Preis eines Heimwerkerartikels in den acht Baumärkten einer Region (in €). Welchen Wert nimmt der Durchschnittspreis an?

k	von über… bis zu …	Einzelwerte	n_k	h_k	$\bar{x}_k$
1	300 – 350	310; 320; 330	3	0,375	$\bar{x}_1 = \frac{1}{3}(310 + 320 + 330) = 320$
2	350 – 400	380; 400	2	0,250	$\bar{x}_2 = \frac{1}{2}(380 + 400) = 390$
3	400 – 450	410; 420; 430	3	0,375	$\bar{x}_3 = \frac{1}{3}(410 + 420 + 430) = 420$
∑			n = 8	1	

Aus den klassierten Daten lässt sich das arithmetische Gesamtmittel berechnen:

- $\bar{x} = \frac{1}{8}\sum_{k=1}^{3} \bar{x}_k\, n_k = \frac{1}{8}(320 \cdot 3 + 390 \cdot 2 + 420 \cdot 3) = \frac{1}{8} \cdot 3000 = 375$
- $\bar{x} = \sum_{k=1}^{3} \bar{x}_k\, h_k = (320 \cdot 0{,}375 + 390 \cdot 0{,}250 + 420 \cdot 0{,}375) = 375$

Man erhält den gleichen Durchschnittswert wie zuvor.

Beispiel: Durchschnittlicher Output

Der Produktionsabteilung liegt der durchschnittliche Output an den unproduktiven Tagen (von über 30 bis zu 50 Teilen), mittelproduktiven Tagen (von über 50 bis zu 70 Teilen) und produktiven Tagen (von über 70 bis zu 100 Teilen) vor. Wie groß ist das arithmetische Mittel insgesamt?

	von über 30 bis zu 50 Teilen	von über 50 bis zu 70 Teilen	von über 70 bis zu 100 Teilen
Durchschnittlicher Output	45	60	90
Anteil der Tage	0,20	0,50	0,30

$$\bar{x} = \sum_{k=1}^{3} \bar{x}_k h_k = 45 \cdot 0{,}20 + 60 \cdot 0{,}50 + 90 \cdot 0{,}30 = 9 + 30 + 27 = 66$$

Berechnung des arithmetischen Mittels bei klassierter Häufigkeitsverteilung (II)

Nutzung der Aggregationseigenschaft: Berechnung des gesamten arithmetischen Mittels durch Aggregation (Zusammenfassung) der arithmetischen Mittel von Teilgesamtheiten.

2. Fall: Unvollständige Information

- Merkmal wurde in klassierter Form erhoben
- Klassierte Häufigkeitsverteilung stammt aus Untersuchungsbericht, in dem nur die Klassengrenzen und die Häufigkeiten verzeichnet sind

→ Klassenmittelwerte $\bar{x}_k$ werden mit Klassenmitten $m_k = \frac{1}{2}(x'_{k-1} + x'_k)$ geschätzt.

Approximatives gesamtes arithmetisches Mittel (Ersetzung der Klassenmittelwerte $\overline{x_k}$ durch die Klassenmitten m_k):

$$\bar{x}_m = \frac{1}{n}\sum_{k=1}^{p} m_k \cdot n_k = \sum_{k=1}^{p} m_k \cdot h_k$$

Beispiel (Fortsetzung): Durchschnittlicher Output

Welchen Gesamtdurchschnittswert würde man für den Output erhalten, wenn die arithmetischen Klassenmittelwerte unbekannt wären?

k	von über… bis zu …	h_k	m_k
1	30 – 50	0,20	$m_1 = \frac{1}{2}(x'_{1-1} + x'_1) = \frac{1}{2}(30 + 50) = 40$
2	50 – 70	0,50	$m_2 = \frac{1}{2}(x'_{2-1} + x'_2) = \frac{1}{2}(50 + 70) = 60$
3	70 – 100	0,30	$m_3 = \frac{1}{2}(x'_{3-1} + x'_3) = \frac{1}{2}(70 + 100) = 85$
∑		1	

$$\bar{x}_m = \sum_{k=1}^{3} m_k \cdot h_k = 40 \cdot 0{,}20 + 60 \cdot 0{,}50 + 85 \cdot 0{,}30 = 8 + 30 + 25{,}5 = 63{,}5$$

Beispiel: Bundesagentur für Arbeit

Eine Erhebung der Bundesagentur für Arbeit hat folgende Verteilung der Arbeitslosendauer ergeben:

Klasse	von über… bis zu …	Anzahl der Arbeitslosen	Prozentsatz der Arbeitslosen
1	0 – 3 Monate	300	30
2	3 – 6 Monate	250	25
3	6 – 12 Monate	250	25
4	12 – 24 Monate	200	20

Gesucht ist die durchschnittliche Arbeitslosendauer. Da keine Klassenmittelwerte bekannt sind, müssen stattdessen die Klassenmitten m_k verwendet werden:

k	von über… bis zu …	n_k	h_k	m_k
1	0 – 3	300	0,30	$^1/_2\,(0+3) = 1{,}5$
2	3 – 6	250	0,25	$^1/_2\,(3+6) = 4{,}5$
3	6 – 12	250	0,25	$^1/_2\,(6+12) = 9$
4	12 – 24	200	0,20	$^1/_2\,(12+24) = 18$
∑		n =1000	1	

Das arithmetische Mittel lässt sich mit Hilfe der absoluten oder relativen Häufigkeiten bestimmen:

$$\bar{x}_m = \frac{1}{n}\sum_{k=1}^{4} m_k n_k = \frac{1}{1000}(1{,}5 \cdot 300 + 4{,}5 \cdot 250 + 9 \cdot 250 + 18 \cdot 200) = 7{,}425$$

$$\bar{x}_m = \sum_{k=1}^{4} m_k h_k = (1{,}5 \cdot 0{,}30 + 4{,}5 \cdot 0{,}25 + 9 \cdot 0{,}25 + 18 \cdot 0{,}20) = 7{,}425$$

Eigenschaften des arithmetischen Mittels (I)

1. Ersatzwerteigenschaft:

   Bei gleichmäßiger Aufteilung der Merkmalssumme erhält jeder Merkmalsträger den Betrag $\bar{x}$:

$$\sum_{i=1}^{n} x_i = n \cdot \bar{x}$$

Beispiel:

In einem Mietshaus verursachten die drei Mieterinnen und Mieter unterschiedlich hohe Heizkosten in einem Monat: 95€, 100€ und 120€. Die Heizkosten werden allerdings nicht separat abgerechnet, sondern auf alle drei Haushalte zu gleichen Teilen umgelegt. Welchen Betrag muss jeder Haushalt bezahlen?

Der zu bezahlende Betrag entspricht dem arithmetischen Mittel:

$$\bar{x} = \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{1}{3}(95 + 100 + 120) = \frac{1}{3} \cdot 315 = 105$$

Eigenschaften des arithmetischen Mittels (II)

2. Schwerpunkteigenschaft:

   Summe der Abweichungen der Einzelwerte vom arithmetischen Mittel ist gleich Null (d.h. positive und negative Abweichungen heben sich gegenseitig auf).

   Bei Einzelwerten:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

   Bei unklassierter Häufigkeitsverteilung:

$$\sum_{j=1}^{m}\left(x_j^* - \bar{x}\right) \cdot n_j = 0$$

Herleitung:

Stellt man die linke Seite von obiger Gleichung als Differenz zweier Summen dar, erhält man:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$$

Da der zweite Term aus n identischen Summanden besteht, lässt er sich durch

$$\sum_{i=1}^{n} \bar{x} = n \cdot \bar{x}$$

vereinfachen, so dass man für

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n \cdot \bar{x}$$

erhält. Unter Verwendung der Ersatzwerteigenschaft folgt unmittelbar die Schwerpunkteigenschaft. Analog kann man die Schwerpunkteigenschaft bei unklassierten Häufigkeitsverteilungen nachweisen.

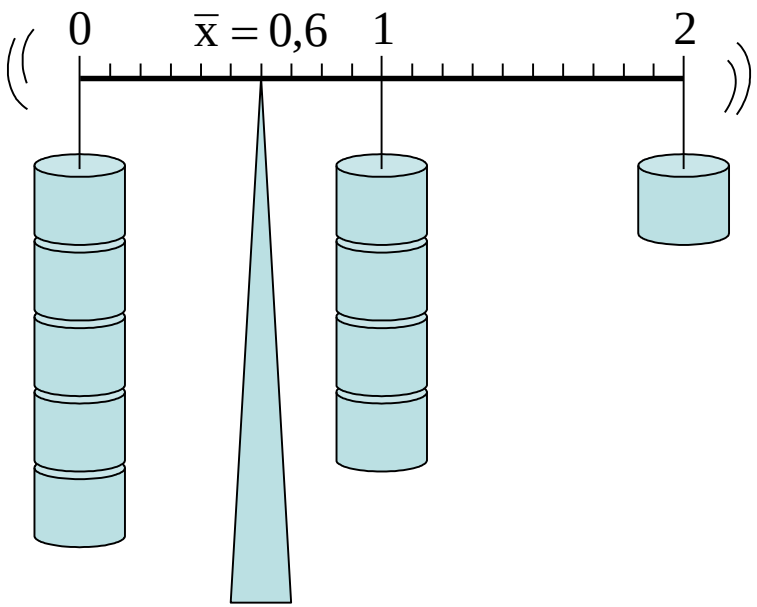
Beispiel (Fortsetzung): Kinderzahl

Die Schwerpunkteigenschaft wird unter Verwendung der Daten zum Merkmal „Kinderzahl“ erläutert. Wenn an einem Balken eine Zahlenskala abgetragen wird und die Gewichte in der Größe von n_j an den Stellen x_j^* befestigt werden, ist der Balken an der Stelle $\bar{x} = 0{,}6$ zu unterstützen, damit er nicht zu einer Seite wegkippt.

Arithmetisches Mittel	Grafische Darstellung

j	x_j^*	n_j	h_j
1	0	5	0,5
2	1	4	0,4
3	2	1	0,1
∑		n = 10	1

$$\bar{x} = \sum_{j=1}^{3} x_j^* \cdot h_j$$
$$= 0 \cdot 0{,}5 + 1 \cdot 0{,}4 + 2 \cdot 0{,}1 = 0{,}6$$



Die Summen der positiven und negativen Abweichungen vom arithmetischen Mittel sind absolut gleich groß, d.h. $(0 - 0{,}6) \cdot 5 = -3$ und $(1 - 0{,}6) \cdot 4 + (2 - 0{,}6) \cdot 1 = 3$.

Eigenschaften des arithmetischen Mittels (III)

3. Lineartransformation:

   In verschiedenen Anwendungen wird von einer Lineartransformation Gebrauch gemacht. Dabei werden alle Einzelwerte um einen konstanten Betrag a und / oder einen multiplikativen Faktor b erhöht.

   Lineartransformation von x:

$$y_i = a + b \cdot x_i$$

   Um das neue arithmetische Mittel von $\bar{y}$ zu bestimmen, müssen nicht alle Einzelwerte neu berechnet werden. Stattdessen kann man das neue arithmetische Mittel durch eine Lineartransformation des alten arithmetischen Mittels $\bar{x}$ ermitteln:

$$\bar{y} = a + b \cdot \bar{x}$$

Beweis:

Als erstes summieren wir beide Seiten der Gleichung $y_i = a + b \cdot x_i$ über alle Merkmalsträger von $i = 1$ bis n auf:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (a + b \cdot x_i)$$

Spaltet man die rechte Seite in zwei Summanden auf, erhält man

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a + \sum_{i=1}^{n} b \cdot x_i = n \cdot a + b \cdot \sum_{i=1}^{n} x_i$$

Nach Division durch n ergibt sich:

$$\frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \cdot n \cdot a + b \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$$\bar{y} = a + b \cdot \bar{x}$$

Beispiel: Lohnerhöhung

Eine Unternehmerin hat drei Angestellte, die 2000€, 2200€, 1500€ verdienen. Bei Tarifverhandlungen wird eine Lohnerhöhung von 3% und ein Sockelbetrag von 50€ pro Monat vereinbart. Den Sockelbetrag erhält jeder Angestellte:

- Der konstante Betrag a, um den alle Einzelwerte ansteigen, beträgt 50.
- Zusätzlich wird der alte Lohn um 3% erhöht. Für den multiplikativen Faktor b ist deshalb 1,03 einzusetzen.

Das neue arithmetische Mittel lässt sich auf zwei Arten bestimmen:

1. Option	2. Option
Neue Einzelwerte: $y_1 = 50 + 1{,}03 \cdot 2000 = 2110$ $y_2 = 50 + 1{,}03 \cdot 2200 = 2316$ $y_3 = 50 + 1{,}03 \cdot 1500 = 1595$ Neues arithmetisches Mittel: $\bar{y} = \frac{1}{3}(2110 + 2316 + 1595) = 2007$	Altes arithmetisches Mittel: $\bar{x} = \frac{1}{3}(2000 + 2200 + 1500) = 1900$ Neues arithmetisches Mittel: $\bar{y} = a + b \cdot \bar{x} = 50 + 1{,}03 \cdot 1900 = 2007$

Eigenschaften des arithmetischen Mittels (IV)

4. Minimumeigenschaft:

   Die Summe der quadrierten Abweichungen einer reellen Zahl c von den Einzelwerten

$$\sum_{i=1}^{n}(x_i - c)^2 \,,$$

   nimmt ein Minimum an, wenn $c = \bar{x}$ ist. Es gilt also

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \leq \sum_{i=1}^{n}(x_i - c)^2 \,.$$

5. Sensitivität gegenüber Ausreißern:

   Folgt unmittelbar aus der Minimumeigenschaft!

Beispiel (Fortsetzung): Umsatzzahlen

Es lagen folgende Umsatzzahlen vor: 3200€, 3000€, 3250€, 3100€, 900€. Der Median sank nur um 50€, als der Ausreißer von 900€ hinzugezogen wurde. Das arithmetische Mittel verringert sich allerdings um mehr als 400€:

Arithmetisches Mittel mit Ausreißer:

$$\bar{x} = \frac{1}{5}(3200 + 3000 + 3250 + 3100 + 900) = 2690$$

Arithmetisches Mittel ohne Ausreißer:

$$\bar{x} = \frac{1}{4}(3200 + 3000 + 3250 + 3100) = 3137$$

→ Liegen Ausreißerwerte vor, dann ist der Median dem arithmetischen Mittel üblicherweise vorzuziehen.

4.4 Geometrisches Mittel

- Anwendung: Hauptsächlich bei Wachstumsvorgängen
- Voraussetzung: Metrisch skaliertes Merkmal

Die durchschnittliche Wachstumsrate ergibt sich durch **geometrische Mittelung** der um eins erhöhten Wachstumsraten (= Wachstumsfaktoren).

- Gegeben: n Einzelwerte eines Merkmals für aufeinander folgende Zeiteinheiten

$$x_1, x_2, \cdots, x_n$$

- Wachstumsrate w_i (prozentuale Interpretation nach Multiplikation mit 100):

$$w_i = \frac{x_i - x_{i-1}}{x_{i-1}}, i = 2, \cdots, n$$

- Wachstumsfaktor f_i:

$$f_i = \frac{x_i}{x_{i-1}} = 1 + w_i, i = 2, \cdots, n$$

- Multipliziert man den (i-1)-ten Einzelwert mit f_i, so erhält man den i-ten Einzelwert:

$$x_i = f_i \cdot x_{i-1}, i = 2, \cdots, n$$

Beispiel: Aktienkurse

Für einen Zeitraum von drei Jahren liegt der Kurs einer Aktie vor. Der Kaufkurs der Aktie beträgt 100€, nach einem Jahr ist sie 120€, nach zwei Jahren 150€ und nach drei Jahren 100€ Wert. Man erhält folgende Wachstumsraten und -faktoren:

i	x_i	w_i (Wachstumsrate)	f_i (Wachstumsfaktor)
1	100	-	-
2	120	$w_2 = \frac{x_2 - x_1}{x_1} = \frac{120 - 100}{100} = 0{,}200\ (20{,}0\%)$	$f_2 = \frac{x_2}{x_{2-1}} = \frac{120}{100} = 1{,}200$ oder $f_2 = 1 + w_2 = 1 + 0{,}200 = 1{,}200$
3	150	$w_3 = \frac{x_3 - x_2}{x_2} = \frac{150 - 120}{120} = 0{,}250\ (25{,}0\%)$	$f_3 = \frac{x_3}{x_{3-1}} = \frac{150}{120} = 1{,}250$ oder $f_3 = 1 + w_3 = 1 + 0{,}250 = 1{,}250$
4	100	$w_4 = \frac{x_4 - x_3}{x_3} = \frac{100 - 150}{150} = -0{,}333\ (-33{,}3\%)$	$f_4 = \frac{x_4}{x_{4-1}} = \frac{100}{150} = 0{,}667$ oder $f_4 = 1 + w_4 = 1 - 0{,}333 = 0{,}667$

Der Wert der Aktie hat sich im ersten Jahr um 20% und im zweiten Jahr um 25% erhöht. Im dritten Jahr ist der Aktienwert um 33,3% gesunken.

Durchschnittliche Wachstumsrate (mit geometrischem Mittel):

$$w = \sqrt[n-1]{f_2 \cdot f_3 \cdot \dots \cdot f_n} - 1 = \sqrt[n-1]{(1+w_2) \cdot (1+w_3) \cdot \dots \cdot (1+w_n)} - 1$$

- Der Wurzelterm kennzeichnet eine geometrische Mittelung
- Der Wert eins muss subtrahiert werden, da Wachstumsfaktoren stets um eins größer sind als Wachstumsraten.
- Es ist die (n-1)-te Wurzel zu ziehen, da es zu n Einzelwerten nur n-1 Wachstumsfaktoren gibt.
- Das **geometrische Mittel** w ist eine durchschnittliche Wachstumsrate. Wird der Einzelwert des Anfangsjahres x_1 (n-1)-mal mit (1+w) multipliziert, dann erhält man den Einzelwert des Endjahres x_n.
- Sind der Anfangswert x_1 und der Endwert x_n bekannt, dann vereinfacht sich die Berechnung des geometrischen Mittels:

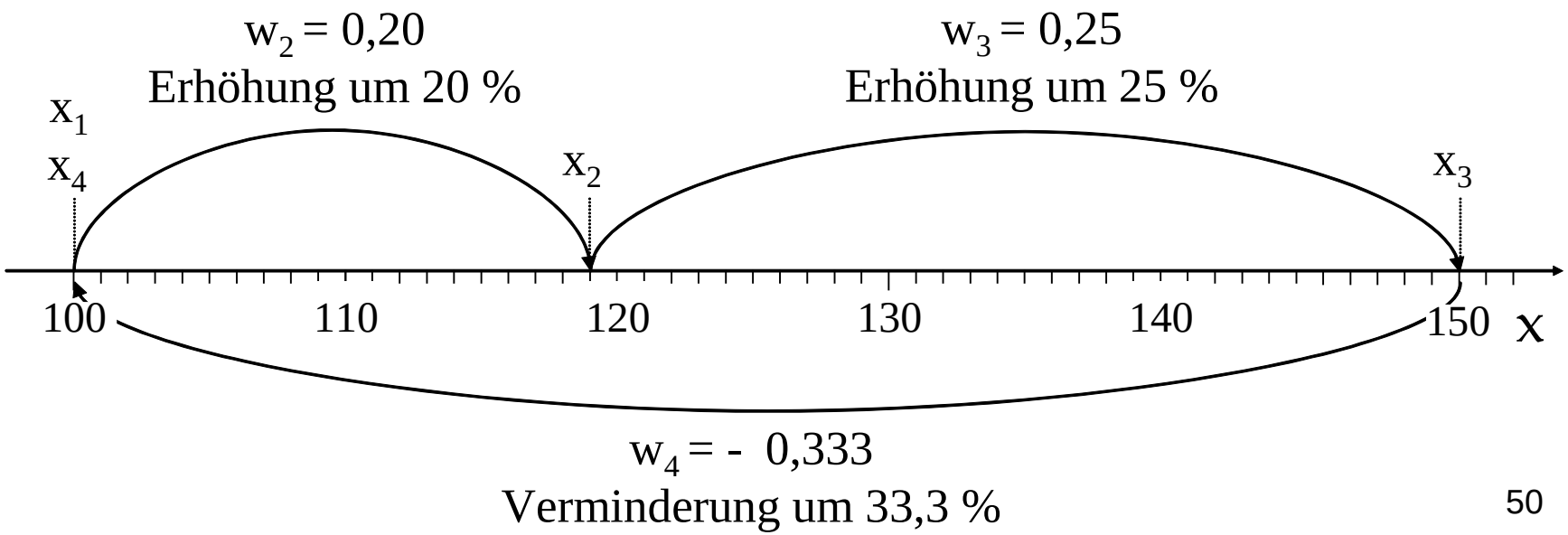
$$w = \sqrt[n-1]{\frac{x_n}{x_1}} - 1 = \left(\frac{x_n}{x_1}\right)^{1/(n-1)} - 1$$

Beispiel (Fortsetzung): Aktienkurse

Wie hoch ist die durchschnittliche Wachstumsrate der Aktienkurse, d.h. die durchschnittliche Rendite? Das arithmetische Mittel ist offenbar nicht sachlich korrekt:

$$\overline{w} = \frac{1}{3}\sum_{i=1}^{3} w_i = \frac{1}{3}(0{,}200 + 0{,}250 - 0{,}333) = \frac{1}{3} \cdot 0{,}117 = 0{,}039 \ (\widehat{=}\ 3{,}9\%)$$

Man würde eine durchschnittliche Erhöhung der Rendite um 3,9 % pro Jahr ausweisen, obwohl sich der Aktienwert über den Gesamtzeitraum nicht verändert hat (Anfangswert und Endwert ist 100€). Die durchschnittliche Wachstumsrate muss also bei 0 liegen.



Beispiel (Fortsetzung): Aktienkurse

Die Anwendung des geometrischen Mittels führt hier zum korrekten Ergebnis, denn eine durchschnittliche Wachstumsrate von Zeitreihenwerten wird immer als geometrisches Mittel bestimmt:

$$\begin{aligned} w &= [(1+w_2)\cdot(1+w_3)\cdot(1+w_4)]^{1/(4-1)} - 1 \\ &= [(1+0{,}200)\cdot(1+0{,}250)\cdot(1-0{,}333)]^{1/(4-1)} - 1 \\ &= [1{,}000]^{1/3} - 1 = 1 - 1 = 0 \ (\widehat{=}\ 0\%) \end{aligned}$$

Oder:

$$w = \left(\frac{x_4}{x_1}\right)^{1/(4-1)} - 1 = \left(\frac{100}{100}\right)^{1/3} - 1 = 1 - 1 = 0$$

Interpretation:

Der Aktienwert hat sich durchschnittlich um 0% pro Jahr verändert. Wird $x_1 = 100$ dreimal mit $1 + w = 1 + 0 = 1$ multipliziert, dann erhält man $x_4 = 100$.

Beispiel: Absatz

Im Basisjahr eins beträgt der Absatz eines Produktes 20.000 Stück und im Jahr sechs 30.000 Stück. Wie hoch ist der durchschnittliche Anstieg des Absatzes?

Ist nach einer durchschnittlichen Veränderung bei Zeitreihenwerten gefragt, dann muss das geometrische Mittel berechnet werden:

$$w = \left(\frac{x_6}{x_1}\right)^{1/(6-1)} - 1 = \left(\frac{30000}{20000}\right)^{1/5} - 1 = 1{,}084 - 1 = 0{,}084\ (\widehat{=}\ 8{,}4\%)$$

Beispiel: Verzinsung

Eine Bank hat folgendes Angebot. Bei einer Laufzeit von 4 Jahren beträgt die Verzinsung des Kapitals im ersten Jahr 3%, im zweiten Jahr 3,5%, im dritten Jahr ebenfalls 3,5% und im vierten Jahr 5%. Die Zinsen verbleiben beim angelegten Geld. Wie hoch ist die durchschnittliche Verzinsung?

Die Zinssätze sind die Wachstumsraten w_i. Die durchschnittliche Verzinsung ist mit dem geometrischen Mittel zu berechnen:

$$\begin{aligned} w &= [(1+w_2)\cdot(1+w_3)\cdot(1+w_4)\cdot(1+w_5)]^{1/(5-1)} - 1 \\ &= [(1+0{,}03)\cdot(1+0{,}035)\cdot(1+0{,}035)\cdot(1+0{,}05)]^{1/(5-1)} - 1 \\ &= 1{,}037 - 1 = 0{,}037 = 0\ (\widehat{=}\ 3{,}7\%) \end{aligned}$$

4.5 Harmonisches Mittel

Das **harmonische Mittel** ist immer dann als durchschnittlicher Wert zu bestimmen, wenn bei einer Verhältniszahl die Zählergröße konstant und die Nennergröße variabel ist.

- Merkmalswerte x_i als Verhältniszahl (Quotienten der Einzelwerte zweier anderer Merkmale A und B):

$$x_i = \frac{a_i}{b_i}$$

- Beispiele für Verhältniszahlen: Weg/Zeit (= Geschwindigkeit), Ausgaben/Preis (= Menge), Kapital/Arbeit (Kapitalintensität), Produktion/Arbeitseinsatz (Arbeitsproduktivität)

- Allgemeine Definition des Durchschnitts bei Verhältniszahlen (Quotient der Summen beider Merkmale), falls Zähler oder Nenner konstant und die andere Größe variabel sind:

$$M = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

Es sind zwei Fälle zu unterscheiden:

Fall 1	**Fall 2**
Harmonisches Mittel: Zählergröße der Verhältniszahl a_i ist konstant ($a_i = a$), Nennergröße b_i ist variabel: $$x_i = \frac{a_i}{b_i} = \frac{a}{b_i} \text{ bzw. } b_i = \frac{a}{x_i}$$ Durch Einsetzen in $$M = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$ erhält man: $$M = \frac{\sum_{i=1}^{n} a}{\sum_{i=1}^{n} \frac{a}{x_i}} = \frac{n \cdot a}{a \cdot \sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = H$$	Ist hingegen die Zählergröße der Verhältniszahl a_i variabel und die Nennergröße b_i aber konstant ($b_i = b$), erhält man zunächst: $$x_i = \frac{a_i}{b_i} = \frac{a_i}{b} \text{ bzw. } a_i = x_i \cdot b$$ Durch Einsetzen in $$M = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$ erhält man das **arithmetische Mittel** $\bar{x}$: $$M = \frac{\sum_{i=1}^{n} x_i \cdot b}{\sum_{i=1}^{n} b} = \frac{b \cdot \sum_{i=1}^{n} x_i}{n \cdot b} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Beispiel: Spedition

Ein Speditionsunternehmen fährt zwei gleich lange Strecken (jeweils 150 km). Die Hinfahrt wird mit einer Geschwindigkeit (= Weg/Zeit) von 150 km/h zurückgelegt, auf der Rückfahrt können aufgrund eines Staus nur 50 km/h gefahren werden. Wie groß ist die Durchschnittsgeschwindigkeit?

Das arithmetische Mittel

$$\bar{x} = \frac{1}{2}\left(150\frac{km}{h} + 50\frac{km}{h}\right) = 100\frac{km}{h}$$

ist offensichtlich sachlich nicht korrekt, weil für die 300 km Strecke dann 3 Stunden benötigen werden. Das Unternehmen ist aber tatsächlich vier Stunden unterwegs, da die Hinstrecke in einer Stunde und die Rückfahrt in drei Stunden zurückgelegt wird. Die Durchschnittsgeschwindigkeit erhält man, indem die gesamte Strecke (150 km + 150 km = 300 km) durch die gesamte Zeit dividiert wird:

$$M = \frac{\sum_{i=1}^{2} a_i}{\sum_{i=1}^{2} b_i} = \frac{150 + 150}{4} = \frac{300}{4} = 75\frac{km}{h} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{2}{\frac{1}{150} + \frac{1}{50}} = H$$

Beispiel: Produktionsstraßen

Ein Unternehmen besitzt drei Produktionsstraßen mit folgender Produktivität:

	1. Produktionsstraße	2. Produktionsstraße	3. Produktionsstraße
Produktivität	5 Stück/h	6 Stück/h	10 Stück/h

An allen Produktionsstraßen werden acht Stunden gearbeitet (konstante Nennergröße). Wie hoch ist die durchschnittliche Produktivität?

Arithmetisches Mittel	Allgemeine Berechnungsformel
Die x-Werte sind die Produktivitäten: $x_1 = 5$, $x_2 = 6$, $x_3 = 10$ Durchschnitt: $\bar{x} = \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{1}{3}(5+6+10) = 7 \left[\frac{Stück}{h}\right]$	Produzierte Stückzahlen in 8 h: $a_1 = 5 \cdot 8 = 40$, $a_2 = 6 \cdot 8 = 48$, $a_3 = 10 \cdot 8 = 80$ Die b_i sind alle 8, weil 8 h an jeder Produktionsstraße gearbeitet werden. Durchschnitt: $M = \frac{\sum_{i=1}^{3} a_i}{\sum_{i=1}^{3} b_i} = \frac{40+48+80}{8+8+8} = 7 \left[\frac{Stück}{h}\right]$

Beispiel (Fortsetzung): Produktionsstraßen

Im Unterschied zu vorher werden jetzt an jeder Produktionsstraße 60 Stück produziert, damit ist die Zählergröße konstant.

Harmonisches Mittel	Allgemeine Berechnungsformel
Die x-Werte sind die Produktivitäten: $x_1 = 5$, $x_2 = 6$, $x_3 = 10$	Die a_i sind alle 60, weil 60 Stück an jeder Produktionsstraße produziert werden. Benötigte Stunden: $b_1 = \frac{60[Stück]}{5\left[Stück/_h\right]} = 12[h]$ $b_2 = \frac{60[Stück]}{6\left[Stück/_h\right]} = 10[h]$ $b_3 = \frac{60[Stück]}{10\left[Stück/_h\right]} = 6[h]$
Durchschnitt: $H = \frac{3}{\sum_{i=1}^{3}\frac{1}{x_i}} = \frac{3}{\frac{1}{5}+\frac{1}{6}+\frac{1}{10}}$ $= 6{,}429\left[\frac{Stück}{h}\right]$	Durchschnitt: $M = \frac{\sum_{i=1}^{3} a_i}{\sum_{i=1}^{3} b_i} = \frac{60+60+60}{12+10+6}$ $= 6{,}429\left[\frac{Stück}{h}\right]$

Berechnung des harmonischen Mittels bei unklassierter Häufigkeitsverteilung:

Sind mehrere Merkmalswerte gleich, dann lässt sich das harmonische Mittel auch mit den Häufigkeiten berechnen:

$$H = \frac{n}{\sum_{j=1}^{m} \frac{1}{x_j^*} \cdot n_j} = \frac{1}{\sum_{j=1}^{m} \frac{1}{x_j^*} \cdot h_j}$$

Beispiel:

Ein Teehändler bestellt monatlich drei Teesorten im Wert von je 1000€. Die ersten beiden Teesorten haben einen Preis (= Ausgaben/Menge) von 2€/100g, für die dritte Teesorte müssen 3€/100g bezahlt werden. Wie hoch ist der Durchschnittspreis?

Da hier die Ausgaben konstant sind (1000€ wird pro Teesorte aufgewendet), ist das harmonische Mittel sachlich korrekt.

- Verwendung der Einzelwerte:

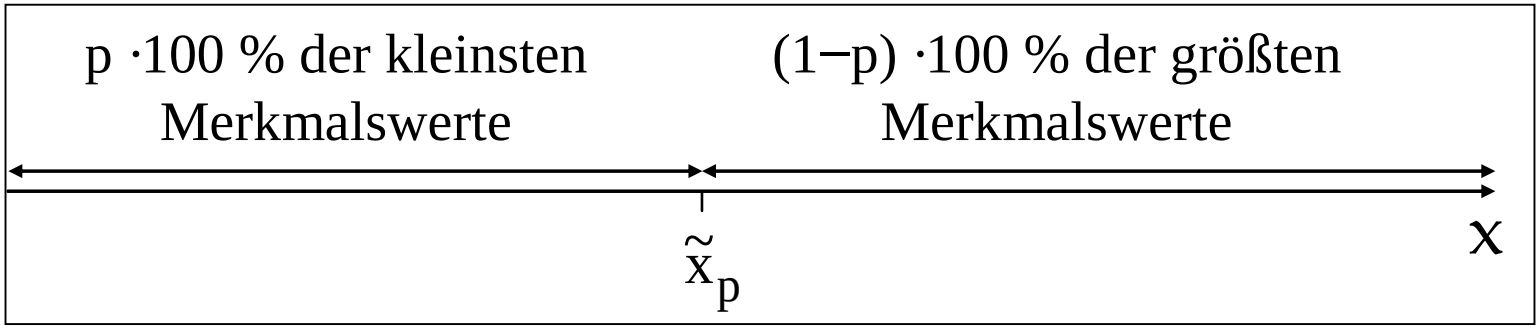
$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{3}{\frac{1}{2} + \frac{1}{2} + \frac{1}{3}} = 2{,}25 \frac{€}{100g}$$

- Verwendung der Häufigkeiten:

$$H = \frac{n}{\sum_{j=1}^{m} \frac{1}{x_j^*} \cdot n_j} = \frac{3}{\frac{1}{2} \cdot 2 + \frac{1}{3} \cdot 1} = 2{,}25 \frac{€}{100g}$$

4.6 Quantile

Das p-Quantil $\tilde{x}_p$ (0 < p <1) trennt die p·100% der kleinsten Merkmalswerte von den übrigen Merkmalswerten.



- Verallgemeinerung des Medians
- Quartile als spezielle Quantile: Vier Intervalle
  - $\tilde{x}_{0,25}$: Das erste Quartil trennt die 25% kleinsten von den 75 % größten Merkmalswerten
  - $\tilde{x}_{0,5} = \tilde{x}$: Das zweite Quartil ist der Median
  - $\tilde{x}_{0,75}$: Mindestens 75% der Merkmalswerte sind kleiner oder gleich dem 3. Quartil und mindestens 25% der Merkmalswerte größer oder gleich dem 3. Quartil

Spezielle Quantile

Art des Quantils	Symbolik	Anzahl der Intervalle
Perzentile	$\tilde{x}_{0,01}, \tilde{x}_{0,02}, \ldots, \tilde{x}_{0,99}$	100
Dezile	$\tilde{x}_{0,1}, \tilde{x}_{0,2}, \ldots, \tilde{x}_{0,9}$	10
Quintile	$\tilde{x}_{0,2}, \tilde{x}_{0,4}, \tilde{x}_{0,6}, \tilde{x}_{0,8}$	5
Quartile	$\tilde{x}_{0,25}, \tilde{x}_{0,5} = \tilde{x}, \tilde{x}_{0,75}$	4
Terzile	$\tilde{x}_{0,333}, \tilde{x}_{0,667}$	3

Berechnung von Quantilen bei Einzelwerten

Gegeben: Geordnete Reihe $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$

$$\tilde{x}_p = \begin{cases} x_{(g+1)} & \text{falls } n \cdot p \text{ nicht ganzzahlig} \\ \frac{1}{2}\left(x_{(g)} + x_{(g+1)}\right) & \text{falls } n \cdot p \text{ ganzzahlig} \end{cases}$$

Dabei ist g der ganzzahlige Teil von n·p, d.h. g = int(n·p).

Die int-Funktion gibt an, den nächstgelegenen kleineren ganzzahligen Wert zu wählen (Abschneiden der Dezimalstellen).

Beispiel (Fortsetzung): Marktforschung

Das zweite Quintil der Heimwerkerartikelpreise (x_1= 310, x_2 = 430, x_3 = 320, x_4 = 330, x_5 = 380, x_6 = 420, x_7 = 400 und x_8 = 410) nimmt folgenden Wert an:

- Geordnete Reihe: $x_{(1)}$ = 310, $x_{(2)}$ = 320,$x_{(3)}$ = 330, $x_{(4)}$ = 380 $x_{(5)}$ = 400, $x_{(6)}$ = 410, $x_{(7)}$ = 420, $x_{(8)}$ = 430
- n·p = 8·2/5 = 8·0,4 = 3,2 (nicht ganzzahlig)
- g = 3 (Nachkommastellen von n·p werden gestrichen)
- $\tilde{x}_{0,4} = x_{(g+1)} = x_{(4)} = 380[€]$

Beispiel (Fortsetzung): Pharmakonzerne

Wie lautet das erste und dritte Quartil der Umsätze der Pharmakonzerne?

- Die aus den Umsätzen (17,3; 16,6; 14,7; 27,0; 17,2; 21,6; 13,6; 40,3; 12,4; 11,7) gebildete geordnete Reihe lautet: $x_{(1)}$ = 11,7; $x_{(2)}$ = 12,4; $x_{(3)}$ = 13,6; $x_{(4)}$ = 14,7; $x_{(5)}$ =16,6; $x_{(6)}$ = 17,2; $x_{(7)}$ = 17,3; $x_{(8)}$ = 21,6; $x_{(9)}$ = 27,0; $x_{(10)}$ = 40,3
- 1. Quartil: n·p = 10·0,25 = 2,5 (nicht ganzzahlig), g = 2 (Nachkommastellen von n·p werden gestrichen) → $\tilde{x}_{0,25} = x_{(g+1)} = x_{(3)} = 13{,}6[\text{Mrd. }\$]$
- 3. Quartil: n·p = 10·0,75 = 7,5 (nicht ganzzahlig), g = 7 (Nachkommastellen von n·p werden gestrichen) → $\tilde{x}_{0,75} = x_{(g+1)} = x_{(8)} = 21{,}6[\text{Mrd. }\$]$

Berechnung von Quantilen bei unklassierter Häufigkeitsverteilung (I)

Verallgemeinerte Median-Formel (0,5 wird durch p ersetzt):

$$\tilde{x}_p = \begin{cases} x_l^* & \text{falls } H_l > p \text{ und } H_{l-1} < p \\ \frac{1}{2}(x_l^* + x_{l+1}^*) & \text{falls } H_l = p \end{cases}$$

Wobei l der Index der Merkmalsausprägung ist, bei der die relativen kumulierten Häufigkeiten p erreichen oder durchlaufen.

Beispiel (Fortsetzung): Beratungsleistungen

Für die Inanspruchnahme von Beratungsleistungen lag folgende Zusammenstellung vor:

Inanspruchnahme von Beratungsleistungen	0	1	2	3
Anteil der Kunden	0,15	0,30	0,35	0,20

Welchen Wert nehmen das erste und dritte Quartil ein?

Berechnung von Quantilen bei unklassierter Häufigkeitsverteilung (II)

Die Kundenanteile stellen die relativen Häufigkeiten dar. Zur Bestimmung des ersten Quartils benötigen wir deren kumulierte Werte, die in einer Tabelle ermittelt werden.

j	x_j^*	h_j	H_j
1	0	0,15	0,15
2	1	0,30	0,15 + 0,30 = 0,45
3	2	0,35	0,45 + 0,35 = 0,80
4	3	0,20	0,80 + 0,20 = 1,0
∑	-	1	-

1. Quartil:

- $l = 2$, weil die relativen kumulierten Häufigkeiten bei der zweiten Merkmalsausprägung $p = 0{,}25$ durchlaufen.
- Da $H_l = H_2 = 0{,}45 > 0{,}25$ und $H_{l-1} = H_1 = 0{,}15 < 0{,}25$ sind, ist keine Mittelung vorzunehmen.
- $\tilde{x}_{0,25} = x_j^* = x_2^* = 1$
- Mindestens 25% der Kunden haben eine oder keine Beratungsleistung in Anspruch genommen, und mindestens 75% eine oder mehr.

Berechnung von Quantilen bei unklassierter Häufigkeitsverteilung (III)

Die Kundenanteile stellen die relativen Häufigkeiten dar. Zur Bestimmung des dritten Quartils benötigen wir deren kumulierte Werte, die in einer Tabelle ermittelt werden.

j	x_j^*	h_j	H_j
1	0	0,15	0,15
2	1	0,30	0,15 + 0,30 = 0,45
3	2	0,35	0,45 + 0,35 = 0,80
4	3	0,20	0,80 + 0,20 = 1,0
∑	-	1	-

3. Quartil:

- $l = 3$, weil die relativen kumulierten Häufigkeiten bei der dritten Merkmalsausprägung $p = 0{,}75$ durchlaufen.
- Da $H_l = H_3 = 0{,}80 > 0{,}75$ und $H_{l-1} = H_2 = 0{,}45 < 0{,}75$ sind, ist keine Mittelung vorzunehmen.
- $\tilde{x}_{0,75} = x_j^* = x_3^* = 2$
- Mindestens 75% der Kunden haben zwei oder weniger Beratungsleistungen in Anspruch genommen, und mindestens 25% zwei oder mehr.

Beispiel (Fortsetzung): Klausurnoten

Berechnet werden soll das zweite Dezil für das Merkmal „Klausurnote“:

j	x_j^*	n_j	h_j	H_j
1	1	1	1/5 = 0,2	0,2
2	3	2	2/5 = 0,4	0,2 + 0,4 = 0,6
3	5	2	2/5 = 0,4	0,6 + 0,4 = 1,0
∑	-	n = 5	1	-

- $l = 1$, weil die relativen kumulierten Häufigkeiten bei der ersten Merkmalsausprägung $p = 0{,}20$ erreichen.
- Da $H_l = H_1 = 0{,}20 = p$, muss gemittelt werden.
- $\tilde{x}_{0,20} = \frac{1}{2}(x_l^* + x_{l+1}^*) = \frac{1}{2}(x_1^* + x_2^*) = \frac{1}{2}(1 + 3) = 2$
- Interpretation: Mindestens 20% der Studierenden haben eine 2 oder eine bessere Note geschrieben, und bei mindestens 80% lag die Klausurnote bei 2 oder schlechter.

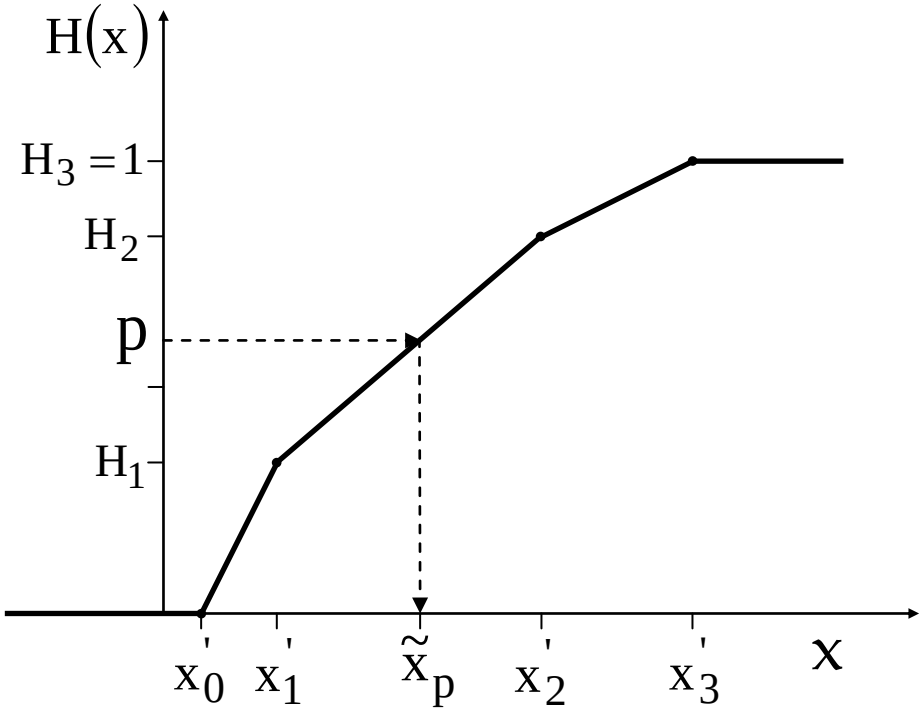
Berechnung von Quantilen bei klassierter Häufigkeitsverteilung

Verallgemeinerte Median-Formel (0,5 wird durch p ersetzt):

$$\tilde{x}_p = x'_{k^*-1} + \frac{p - H_{k^*-1}}{h_{k^*}} \cdot b_{k^*}$$

kobei k* die Klasse ist, bei der die H_k p durchlaufen.

Bestimmung des p-Quantils durch Interpolation:



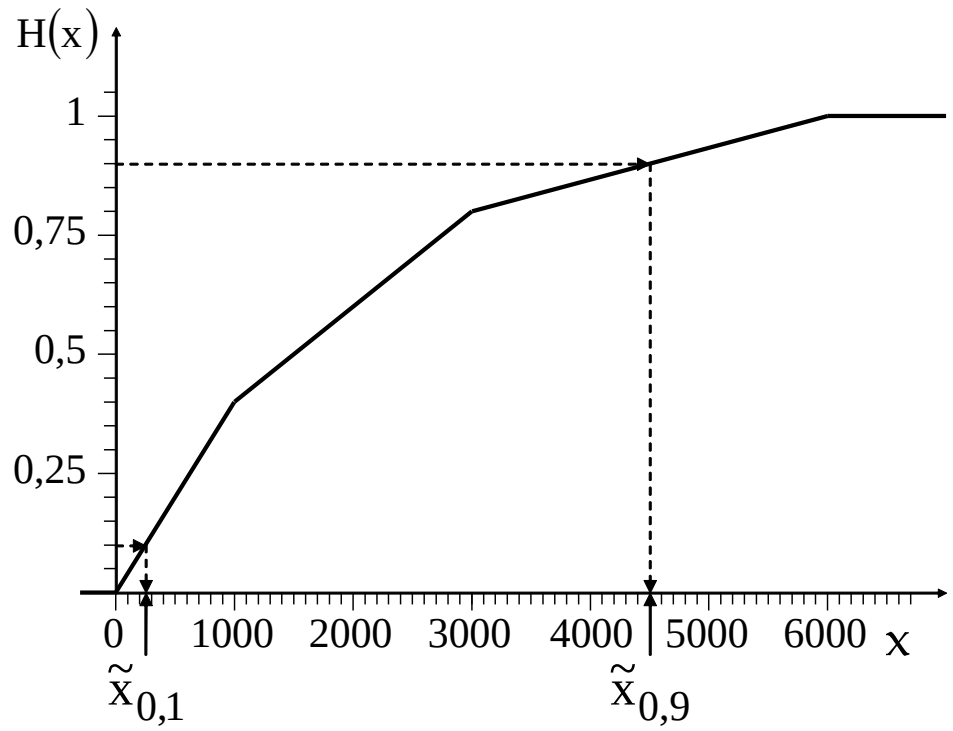
Beispiel (Forsetzung): Einkommensverteilung

In der Diskussion über die Einkommensverteilung interessiert, welchen Betrag die 10% der Bezieher der geringsten Einkommen (1. Dezil) höchstens und die 10% der Bezieher der höchsten Einkommen (9. Dezil) mindestens verdienen.

Kumulierte relative Häufigkeiten

k	von über… bis zu …	h_k	H_k
1	0 – 1000	0,4	0,4
2	1000 – 3000	0,4	0,8
3	3000 – 6000	0,2	1
∑		1	

Grafische Darstellung



Beispiel (Fortsetzung): Einkommensverteilung

1. Dezil:

- $k^* = 1$, da die H_k in der ersten Klasse $p = 0{,}1$ durchlaufen.
- $\tilde{x}_{0,1} = x'_{k^*-1} + \frac{0{,}1 - H_{k^*-1}}{h_{k^*}} \cdot b_{k^*} = x'_0 + \frac{0{,}1 - H_0}{h_1} \cdot b_1 = 0 + \frac{0{,}1 - 0}{0{,}4} \cdot 1000 = 250[€]$

9. Dezil:

- $k^* = 3$, da die H_k in der dritten Klasse $p = 0{,}9$ durchlaufen.
- $\tilde{x}_{0,9} = x'_{k^*-1} + \frac{0{,}9 - H_{k^*-1}}{h_{k^*}} \cdot b_{k^*} = x'_2 + \frac{0{,}9 - H_2}{h_3} \cdot b_3 = 3000 + \frac{0{,}9 - 0{,}8}{0{,}2} \cdot 3000 = 4500[€]$

Somit verdienen (näherungsweise) 10% der Befragten bis zu 250 €, während ebenso 10% ein Einkommen von mindestens 4500 € erzielen.