Betriebsstatistik im Industrieunternehmen

Wolf-Dieter Rückwart

1 Basis der Betriebsstatistik

Die Statistik im Industriebetrieb befasst sich mit dem Sammeln, Aufbereiten und Auswerten von Mengen- und Wertgrößen, die für die Überwachung des Betriebsgeschehens sowie für die Vorbereitung unternehmerischer Entscheidungen wichtig sind. Hierzu

- werden die im Unternehmen anfallenden Daten nach **Art** sowie nach **Menge** oder **Wert** geordnet gesammelt.
- werden die gesammelten Daten (Urdaten) nach **Häufigkeit** und/oder **Klassierung** zu geeigneten **Tabellen** und **Grafiken** verkürzt.
- werden mit Hilfe statischer Verfahren aus den Tabellen und Grafiken **Kennzahlen** (Mittelwerte, Streuungsmaße) gewonnen und als Grundlage für **Kontrollen** und **Dispositionen** an die Unternehmensleitung weitergeleitet.

Gegenstand statistischer Betrachtungen sind nicht Einzelerscheinungen, sondern wiederkehrende Ereignisse oder Veränderungen, die sich entweder auf einen bestimmten Zeitpunkt oder einen bestimmten Zeitraum beziehen; danach unterscheidet man:

Zeitpunktbezogene Statistik	Zeitraumbezogene Statistik
Bilanzanalyse	Beschaffungsstatistik
Analyse der Gewinn- und Verlustrechnung	Absatzstatistik
Lagerstatistik	Personalstatistik

2 Tabellen in der Betriebsstatistik

Die wichtigsten innerbetrieblichen Quellen, aus denen die Betriebsstatistik ihre Zahlen gewinnt, sind die Personalabteilung, die Finanzbuchhaltung sowie die Kosten- und Leistungsrechnung . Die dort erfassten Größen stellen die **Urdaten** dar, die sich wegen ihrer Fülle nicht für eine Betriebskontrolle oder zur Vorbereitung unternehmerischer Entscheidungen eigenen. Sie gewinnen ihre statistische Qualität erst, wenn sie in einer verkürzten **Tabelle** folgende Angaben enthalten:

- einen **Merkmalsträger,**
- ein bestimmtes **Merkmal**,
- bestimmte **Merkmalsausprägungen**

Ausgangsbeispiel

In der Thomas Berg KG sind zum Zeitpunkt 01.01.01 insgesamt 25 Arbeitnehmer im kaufmännischen und gewerblichen Bereich beschäftigt. Aus den vorliegenden Personaldaten soll die Altersstruktur für eine angedachte Betriebsrente (und der damit verbundenen Aufwendungen) anonym ermittelt werden. Die Sachbearbeiterin stellt folgende Altersangaben in Jahren in alphabetischer Reihenfolge der Mitarbeiternamen zusammen, bereitet die Zahlen statistisch auf und leitet die Ergebnisse an Thomas Berg weiter:
23, 35, 38, 25, 23, 35, 56, 23, 25, 42, 59, 38, 38, 42, 35, 59, 56, 23, 48, 42, 47, 25, 48, 29, 42

Typische und markante Erscheinungen lassen sich erst durch Zusammenfassung gleichartiger Größen aus den **Urdaten** nach bestimmten Merkmalsausprägungen erkennen. Im obigen Beispiel sind die **Arbeitnehmer** und **Arbeitnehmerinnen** der Thomas Berg KG die **Merkmalsträger**. Sie sollen nach dem **Merkmal „Altersstruktur“** untersucht werden. Hierzu dienen als **Merkmalsausprägungen** die Angaben der **Lebensalter** in Jahren (von 23 bis 59 Jahre).

Fortsetzung des Beispiels:
Als erstes Ergebnis erstellt die Sachbearbeiterin eine noch unübersichtliche **Tabelle**, in der sie das Alter der Arbeitnehmer in eine aufsteigende Reihenfolge bringt.

Tabelle: Alter der Arbeitnehmer zum 01.01.01

Arbeitnehmer (laufende Nummer)	Alter der Arbeitnehmer in Jahren in aufsteigender Reihenfolge
1	23
2	23
3	23
4	23
5	25
6	25
7	25
8	29
9	35
10	35
11	35
12	38
13	38
14	38
15	42
16	42
17	42
18	42
19	47
20	48
21	48
22	56
23	56
24	59
25	59

Die obige Tabelle besagt lediglich, dass 25 Arbeitnehmer vorhanden sind, von denen mehrere im gleichen Alter sind; die jüngsten Arbeitnehmer sind 23, die ältesten 59. Erst die Aufbereitung der Daten mit Blick auf eine Untersuchungsfrage schafft eine sinnvolle statistische Grundlage, insbesondere wenn die Urdaten deutlich mehr Merkmalsausprägungen enthalten.

Eine statistische Tabelle ist durch die Zahlenanordnung in Spalten und Zeilen gekennzeichnet. Den Namen der Tabelle erläutert der sog. Tabellenkopf. Der Inhalt wird durch die sog. Vorspalte benannt; danach folgen die eigentlichen Tabellenspalten. Der Platz, der für die einzelne Eintragung in den Tabellenspalten vorgesehen ist, heißt Tabellenfach (Tabellenfeld).

Ihre Aufgabe erfüllen Tabellen nur dann, wenn bei ihrer Erstellung die folgenden wesentlichen Gesichtspunkte beachtet werden:

- klare Überschriften im Tabellenkopf und in der Vorspalte,
- möglichst wenige Einteilungsmerkmale, damit die Übersichtlichkeit gewahrt bleibt,
- zweckmäßiger Aufbau, damit das Lesen der Tabelle erleichtert wird,

- Ausrichtung auf eine Untersuchungsfrage.

2.1 Häufigkeitsverteilung

Es liegt nahe, die obige Tabelle aufgrund der mehrfach vorkommenden gleichen Altersstufe über die sog. **Häufigkeitsverteilung** zu verkürzen. Eine Häufigkeitsverteilung liegt dann vor, wenn eine oder mehrere Merkmalsausprägungen (hier: Alter) mehrfach vorkommen. Die Merkmalsausprägungen rücken mit einmaliger Nennung tabellarisch in die Vorspalte; das folgende Tabellenfach nimmt die Häufigkeiten auf, mit der die jeweilige Merkmalsausprägung vertreten ist. Auf diese Weise entstehend eine verkürzte Tabelle der **absoluten Häufigkeiten**, die noch durch sog. **relative Häufigkeiten** ergänzt werden kann. Relative Häufigkeiten sind nichts anderes als die in Prozentzahlen ausgedrückten absoluten Häufigkeiten. Zweckmäßig ist es, die relativen Häufigkeiten schrittweise aufzusummieren (**aufsummierte Häufigkeit**). Mit den so gewonnenen Daten sind erste Interpretationen hinsichtlich einer Untersuchungsfrage – z. B. Altersstruktur - möglich.

Fortsetzung des Beispiels

Tabelle: Alter der Arbeitnehmer zum 01.01.01 nach Häufigkeiten

Merkmalsausprägungen (Alter)	Absolute Häufigkeit n_i	Relative Häufigkeit f_i in %	Aufsummierte Häufigkeit in %
23	4	16,0 %	16,0 %
25	3	12,0 %	28,0 %
29	1	4,0 %	32,0 %
35	3	12,0 %	44,0 %
38	3	12,0 %	56,0 %
42	4	16,0 %	72,0 %
47	1	4,0 %	76,0 %
48	2	8,0 %	84,0 %
56	2	8,0 %	92,0 %
59	2	8,0 %	100,0 %
--	n = 25	100 %	--

Die relative Häufigkeit wird mit $\mathbf{f_i}$ bezeichnet, die Summe der untersuchten Merkmalsträger mit **n** und die Häufigkeit, mit der die Merkmalträger vorkommen, mit $\mathbf{n_i}$. Die Rechenvorschrift zur Berechnung der relativen Häufigkeit lautet also:

$\mathbf{f_i = n_i : n}$

Die relative Häufigkeit f_i beträgt dann im ersten Fall für das Alter von 23 Jahren:

$f_i = 4 : 25 = 0{,}16 = 16\ \%$

Entsprechend werden die übrigen Prozentzahlen und die aufsummierte Häufigkeit bestimmt.

Die Tabelle enthält nur noch 10 Altersangaben mit den zugehörigen Häufigkeiten. Aus ihr ist mit Hilfe der relativen Häufigkeit abzulesen, dass z. B. die älteste Gruppe der Arbeitnehmer (59 Jahre) nur 8 % der Belegschaft ausmacht. Die aufsummierte Häufigkeit macht u. a. darauf aufmerksam, dass mehr als die Hälfte der Arbeitnehmer (56,0 %) im Alter zwischen 23 und 38 Jahren ist.

Eine grafische Darstellung der Tabellenergebnisse, die ein anschauliches Bild der Tabelle vermittelt, ist hier nur mit Lücken in den Merkmalsausprägungen (Alter) möglich. Die Lücken verhindern ein kontinuierlich verlaufendes Diagramm. Erst über die sog. **Klassenbildung** verläuft die Betrachtung in einer Tabelle oder einer Graphik kontinuierlich.

2.2 Klassenbildung

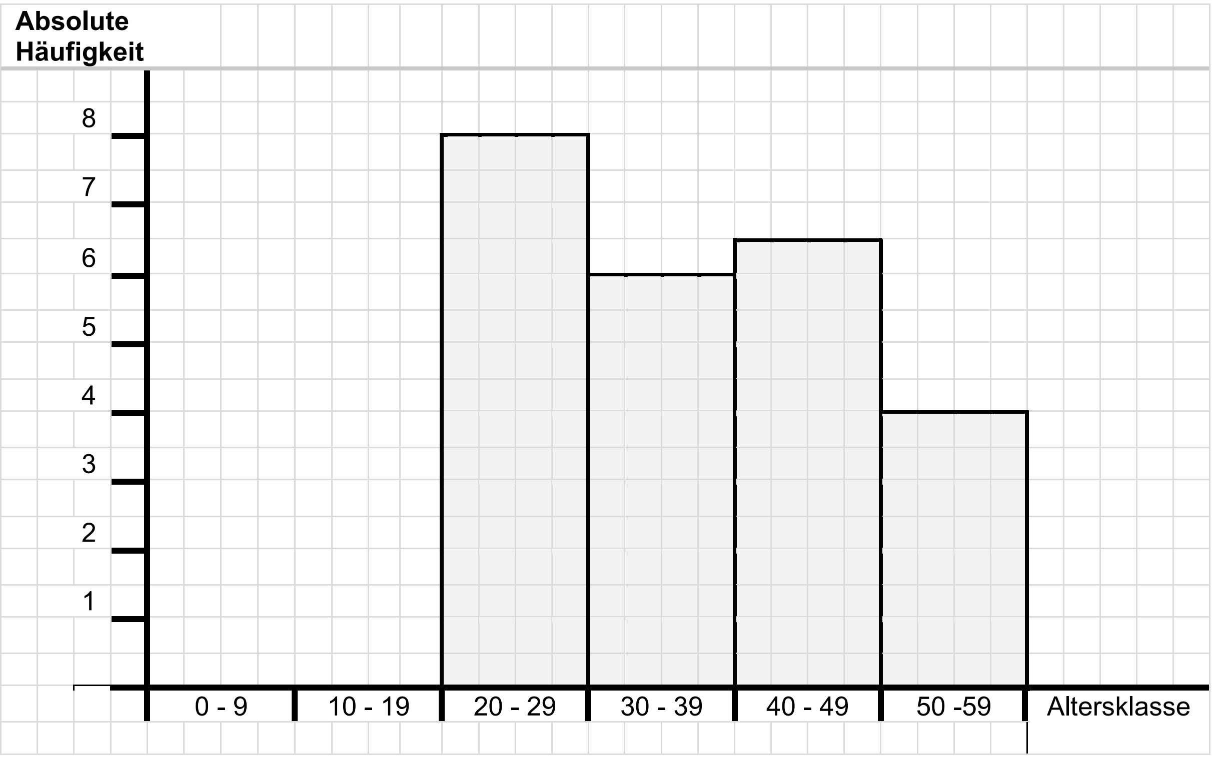
Bei der Vielzahl an Merkmalsausprägungen im Beispiel liegt es nahe, die Anzahl der Ausprägungen durch eine Klassenbildung überschaubar zu machen. Im Beispiel bietet sich an, Altersgruppen für jedes Jahrzehnt zu bilden. Die Klassenbreite würde dann also 10 Jahre mit entsprechenden absoluten und relativen Häufigkeiten betragen. Die Klassenbildung unterstützt zudem die Untersuchungsfrage.

Fortsetzung des Beispiels

Tabelle: Altersklassen der Arbeitnehmer zum 01.01.01 nach Häufigkeiten

Altersklassen in Jahren	Absolute Häufigkeit	Relative Häufigkeit f_i in %	Aufsummierte Häufigkeit in %
20 bis unter 30 alternativ [20;30)[1]	8	32,0 %	32,0 %
30 bis unter 40 alternativ [30;40)	6	24,0 %	56,0 %
40 bis unter 50 alternativ [40;50)	7	28,0 %	84,0 %
50 bis unter 60 alternativ [50;60)	4	16,0 %	100,0 %
	25	100,0 %	--

Alle Klassen haben die gleiche Breite von 10 Jahren. Durch die Klassenbildung wird die Tabelle verkürzt. Zum Nachteil gereicht, dass dadurch Einzelheiten verloren gehen, andererseits wird die Lesbarkeit und die Übersichtlichkeit im Sinne der Untersuchungsfrage „Altersstruktur "erhöht. Zudem kann in diesem Fall die graphische Darstellung der **absoluten Häufigkeit** über ein **Säulendiagramm** die Anschaulichkeit unterstützen:

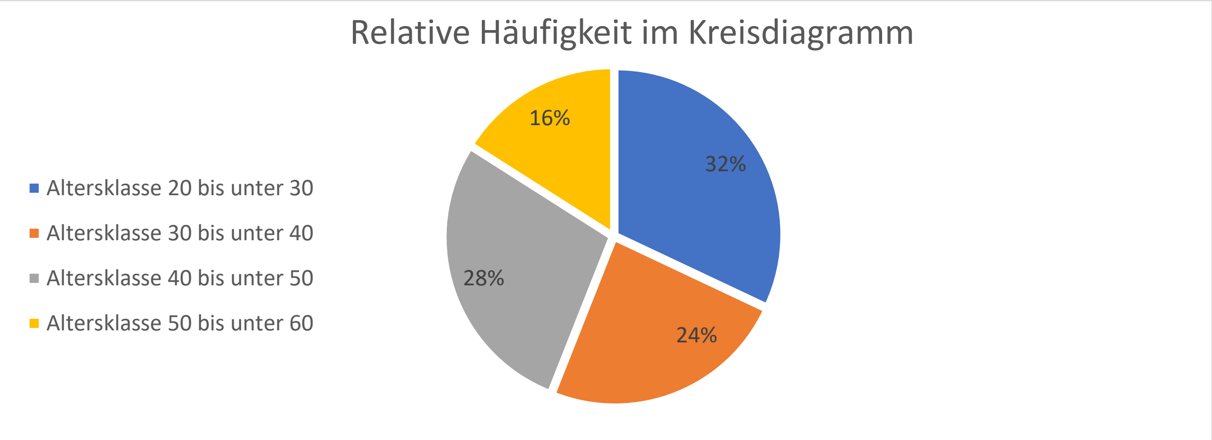


[1] Durch die Schreibweise [a;b) wird festgelegt, dass die Zahl der unteren Grenze eingeschlossen ist, während die Zahl der obere Grenze ausgeschlossen wird.

Die absolute Häufigkeit zeigt, dass die Altersklasse „20 bis 29" mit der höchsten Häufigkeit vertreten ist und dass mit zunehmendem Alter – insbesondere in der höchsten Altersklasse – die Häufigkeit abnimmt.

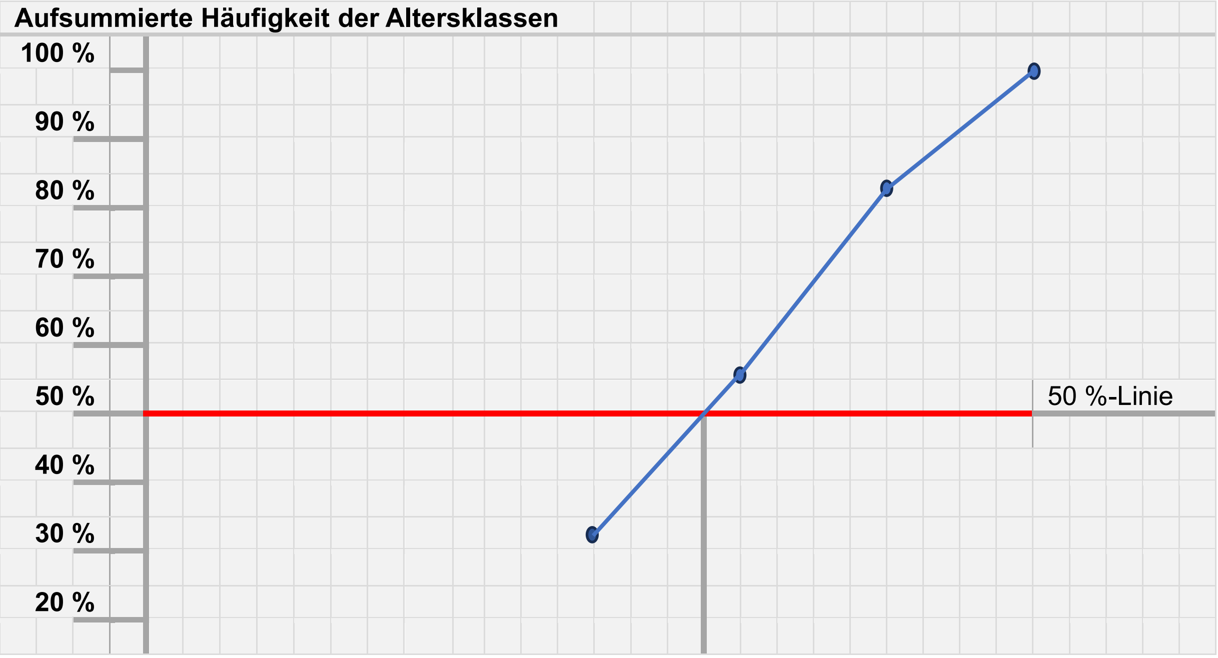
Es bietet sich an, die **relative Häufigkeit** in Prozent anzugeben, da somit wegen der geringen Anzahl von Klasseneinteilungen die Darstellung in einem **Kreisdiagramm** möglich ist. Die gesamte Kreisfläche (≙ 360°) entspricht 100 %. Für die einzelnen relativen Häufigkeiten sind über die Winkelgrade die zugehörigen Kreissektoren zu ermitteln:

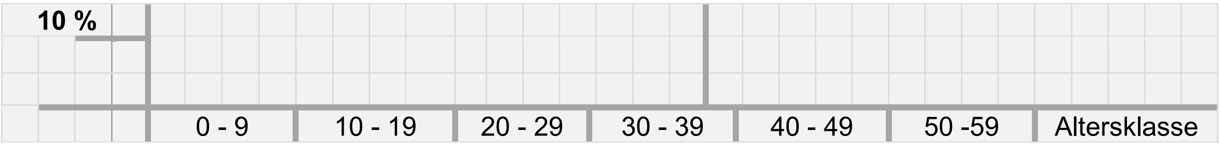
Altersklassen	Relative Häufigkeit in %	Kreissektoren in Grad
20 bis unter 30	32 %	115 °
30 bis unter 40	24 %	86 °
40 bis unter 50	28 %	101 °
50 bis unter 60	16 %	58 °
	100 %	360 °



Das Kreisdiagramm veranschaulicht auf übersichtliche Weise, dass u. a. über die Hälfte der Arbeitnehmer in den Altersbereich „20 bis 40 Jahre" fallen.

Die **aufsummierte Häufigkeit** unterstützt statistisch die zu untersuchende Frage nach der Altersstruktur. Sie kann in einem **Liniendiagramm** verdeutlicht werden, in das die 50 %-Linie eingezeichnet ist. Die Häufigkeiten werden als Punkte jeweils am Ende der Altersklasse angesetzt und die Punkte durch eine Linie verbunden.





Die 50 %-Linie der Altersstruktur lässt deutlicher als die Darstellung der absoluten Häufigkeit erkennen, dass die Altersklasse bis 39 Jahren besonders stark vertreten ist.

3 Mittelwerte

Im Kapitel 2 haben Sie erfahren, dass durch Häufigkeiten und Klassenbildung ein ungeordnetes Datenmaterial übersichtlich gestaltet werden kann, um daraus erste Auswertungen zu gewinnen. Die fundiertere Interpretation der Eigenschaften von Datensätzen (Tabellen) mit oder ohne Häufigkeitsverteilungen ist über die Mittelwerte möglich. Zu ihnen zählen u. a. neben dem arithmetischen Mittel auch der Zentralwert (Median).

Mittelwerte sind in der Statistik charakteristische Stellvertreter für viele Merkmalsausprägungen des gleichen Merkmals. Sie geben die zentrale Tendenz der Merkmalsausprägungen an und eignen sich für inner- und zwischenbetriebliche Vergleiche.

3.1 Arithmetisches Mittel

Beim arithmetischen Mittel – ohne Berücksichtigung der Häufigkeiten - werden die einzelnen Daten einer Stichprobe oder des Urmaterials aufaddiert und durch die Anzahl der Werte dividiert. Arithmetische Mittel finden z. B. in der Personal- und Bilanzauswertung sowie in der Bewertung des Vorratsvermögens und bei Mischungsverhältnissen Anwendung.

Fortsetzung des Beispiels:
Aus den ursprünglichen Altersdaten von Seite 1 soll das arithmetische Mittel als durchschnittliches Alter ermittelt werden:
23, 35, 38, 25, 23, 35, 56, 23, 25, 42, 59, 38, 38, 42, 35, 59, 56, 23, 48, 42, 47, 25, 48, 29, 42

In der Statistik wird das arithmetische Mittel mit dem Symbol $\bar{x}$ bezeichnet.

$$\bar{x} = \frac{23+35+38+25+23+35+56+23+25+42+59+38+38+42+35+59+56+23+48+42+47+25+48+29+42}{25}$$

$$\bar{x} = \frac{956}{25} = \mathbf{38{,}24}\text{ Jahre}$$

Die Arbeitnehmer der Thomas Berg KG sind im Durchschnitt 38,24 Jahre alt. Sie sind damit relativ jung und Thomas Berg muss für die angedachte Betriebsrente erhebliche Aufwendungen tätigen.

Allgemein wird die Anzahl der Merkmalsauspägungen mit **n** bezeichnet, die Merkmalsausprägungen selbst mit $\mathbf{x_1, x_2, x_3, \ldots x_i, \ldots, x_n}$. Das ergibt sich die allgemeine Gleichung zur Berechnung des arithmetischen Mittels:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Die Gleichung lässt sich auch mit Hilfe des Summenzeichens wie folgt schreiben:

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

Bei einer hohen Anzahl der Merkmalsausprägungen wird die Berechnung des Mittelwertes sehr aufwändig. Übersichtlicher wird die Berechnung über das gewogene arithmetische Mittel.

3.2 Gewogenes arithmetisches Mittel

Das gewogene arithmetische Mittel berücksichtigt die Häufigkeit, in der die Merkmalausprägungen vorkommen.

Aus der Häufigkeit multipliziert mit der jeweiligen Merkmalsausprägung ergibt sich in verkürzter Form die gleiche Summe wie im vorhergehenden Beispiel. Diese Summe ist durch die Anzahl der Merkmalausprägungen (Häufigkeiten) zu dividieren.

Fortsetzung des Beispiels:
Das Beispiel der Häufigkeitsverteilung von Seite 3 ist die Grundlage für das gewogene arithmetische Mittel.

$$\bar{x} = \frac{23 \cdot 4 + 25 \cdot 3 + 29 + 35 \cdot 3 + 38 \cdot 3 + 42 \cdot 4 + 47 + 48 \cdot 2 + 56 \cdot 2 + 59 \cdot 2}{25}$$

$$\bar{\mathbf{x}} = \frac{956}{25} = \mathbf{38{,}24} \text{ Jahre}$$

Das Ergebnis von 38,24 Jahren ist auch bei verkürzter Berechnung unter Beachtung der Häufigkeitsverteilung das gleiche wie zuvor.

Allgemein wird hier die Häufigkeit mit $\mathbf{n_j}$ (mit $1 \leq j \leq m$) bezeichnet, die Merkmalsausprägungen mit $\mathbf{x_1, x_2, x_3, \ldots, x_j, \ldots x_m}$. Das ergibt sich die allgemeine Gleichung zur Berechnung des gewogenen arithmetischen Mittels:

$$\bar{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 + \ldots + n_m \cdot x_m}{n}$$

$$\bar{x} = \frac{1}{n} \cdot \sum_{j=1}^{m} x_j \cdot n_j$$

Zu einem anderen Ergebnis gelangt die Berechnung mit Hilfe der Klassenbildung, wobei hier die „Ungenauigkeit“ besteht, die Klassenmitten als repräsentativ für die Merkmalausprägungen anzunehmen. Die Klassenbildung führt zu Informationsverlusten.

Fortsetzung des Beispiels:
Das Beispiel der Klasseneinteilung von Seite 4 ist die Grundlage für das gewogene arithmetische Mittel.

Die Klassenmitten haben die Werte 25, 35, 45 und 55; daraus ergibt sich:

$$\bar{x} = \frac{25 \cdot 8 + 6 \cdot 35 + 7 \cdot 45 + 4 \cdot 55}{25}$$

$$\bar{x} = \underline{200 + 210 + 315 + 220}$$

$$\bar{X} = \frac{945}{25} = 37{,}8 \text{ Jahre}$$

Die Abweichung gegenüber dem tatsächlichen Durchschnittalter (38,24 Jahre) beträgt 0,44 Jahre.

Fazit zum arithmetischen Mittel
Mit der Berechnung des Durchschnittsalters wird eine erste quantitative Auswertung geschaffen. Bei relativ gleichmäßiger Verteilung der Merkmalsausprägungen über die Spannweite des Alters (23 Jahre bis 59) ist das arithmetische Mittel ein guter Beurteilungsmaßstab. In Verbindung mit der Tabelle und den grafischen Darstellungen zur relativen Häufigkeit und zur aufsummierten Häufigkeit liefert es dem Betrachter ein zutreffendes Bild der Altersstruktur.

Das arithmetische Mittel ist bei Extremwerten in der Urdatei nur begrenzt brauchbar.

Beispiel:

Die Gehälter in einem Unternehmen weisen neben den tariflich gebundenen Entgelten, die relativ nahe beieinander liegen, auch Entgelte für die Geschäftsleitung auf, die sich deutlich von den tariflich gebundenen Gehältern abheben. Die geordnete Auflistung der Gehälter ergibt folgende Zahlenreihe:

2.780,00 €, 3.135,00 €, 3.275,00 €, 3.560,00 €, 3.615,00 €, 4.065,00 €, 4.135,00 €, 8.155,00 €, 9.240,00 €.

Das arithmetische Mittel beträgt in diesem Fall 41.960,00 € **:** 9 = **4.662,22 €**
Geeigneter als das arithmetische Mittel wäre hier der Zentralwert (3.615,00 €, s. Seite 10).

3.3 Zentralwert

Der Zentralwert, auch Median genannt, bezeichnet die Mitte einer geordneten Datenreihe.

Fortsetzung des Beispiels:

Im Beispiel auf Seite 2 sind die Arbeitnehmer der Thomas Berg KG nach ihrem Alter geordnet; insgesamt umfasst die Datenreihe 25 Werte, also eine **ungerade** Anzahl. Die Mitte davon liegt auf der 13. Merkmalsausprägung (38 Jahre). In diesem Fall liegen 12 Werte oberhalb des Zentralwertes und 12 darunter:

Tabelle: Geordnete Altersangabe der Arbeitnehmer mit Zentralwert

Laufende Nummer	Alter der Arbeitnehmer in Jahren
1	23
2	23
3	23
4	23
5	25
6	25
7	25
8	29
9	35
10	35
11	35
12	38
13 = **(n+1)/2**	**Zentralwert (Z)** → **38**

14	38
15	42
16	42
17	42
18	42
19	47
20	48
21	48
22	56
23	56
24	59
25 = n	59

Der Zentralwert Z wird in diesem Fall berechnet, indem man den Umfang der Datenreihe (25) um 1 erhöht und die Summe durch 2 dividiert. Anschließend ist die zugehörige Merkmalsausprägung aufzusuchen:

(25 + 1) **:** 2 = 13
Zur 13. Ziffer auf der Zahlenreihe gehört der Zentralwert **Z = 38**

Allgemein:
Der **Zentralwert Z** wird bei ungeraden Datenreihen berechnet, indem aus **(n + 1) : 2** die mittlere Position bestimmt wird; die zugehörige Merkmalsausprägung ist der Zentralwert.

Der Zentralwert unterstützt in diesem Fall die zuvor gemachte Aussage nach dem Durchschnittsalter auf der Basis des arithmetischen Mittels (38,24 Jahre).

Ist die Anzahl der Daten in einer Datenreihe eine **gerade** Zahl, so liegt der Zentralwert zwischen den beiden mittleren Werten. Der niedrigere Wert liegt an der Position **n/2**, der höhere an der Position (**n/2 + 1)**. Der Zentralwert ist dann der Mittelwert aus beiden.

Fortsetzung des Beispiels:
Das obige Beispiel wird so abgeändert, dass der letzte Wert entfällt. Es bleiben dann 24 Werte übrig, also eine gerade Anzahl.

Tabelle: Geordnete Altersangabe der Arbeitnehmer mit Zentralwert

Laufende Nummer	Alter der Arbeitnehmer in Jahren
1	23
2	23
3	23
4	23
5	25
6	25
7	25
8	29
9	35
10	35
11	35
12 = n/2	Zentralwert → 38
13 = n/2 + 1	38 **38**
14	38
15	42
16	42
17	42
18	42
19	47
20	48
21	48
22	56

23	56
24 = n	59

Berechnung des Zentralwertes bei geraden Merkmalsausprägungen:
(1) (24/2 + 24/2 + 1) **:** 2 = 12,5
(2) **Z** = (38 + 38) **:** 2 = **38**

Fazit zum Zentralwert
Der Zentralwert eignet sich zur Auswertung einer statistischen Datenreihe insbesondere dann, wenn innerhalb der Datenreihe extreme Werte vorkommen. Im obigen Beispiel ist der Zentralwert (38 Jahre) kaum vom arithmetischen Mittel (38,24 Jahre) zu unterscheiden. In diesem Fall würde das arithmetische Mittel zur Auswertung ausreichen. Im Beispiel der Gehaltsreihe von Seite 8 ist der Zentralwert geeigneter als das arithmetische Mittel, da er die extremen Werte nicht beachtet.

Beispiel (vgl. Seite 8)**:**

Die Gehälter in einem Unternehmen weisen neben den tariflich gebundenen Entgelten, die relativ nahe beieinander liegen, auch Entgelte für die Geschäftsleitung auf, die sich deutlich von den tariflich gebundenen Gehältern abheben. Die geordnete Auflistung der Gehälter ergibt folgende Zahlenreihe:

2.780,00 €, 3.135,00 €, 3.275,00 €, 3.560,00 €, **3.615,00 €**, 4.065,00 €, 4.135,00 €, 8.155,00 €, 9.240,00 €.

Zentralwert

Der Zentralwert beträgt im obigen Gehaltsbeispiel als 5. Wert:
Z = 3.615,00 €
Er unterscheidet sich damit deutlich vom arithmetischen Mittel und gibt die Datenreihe statistisch gut wieder.

4 Streuungsmaße

Streuungsmaße werden in der Regel als Abweichungen aus den beobachteten Mittelwerten berechnet. Wenn sich das Streuungsmaß eng an den Mittelwert anschmiegt, repräsentiert ein Mittelwert die Merkmalsausprägungen einer Datenreihe hinreichend gut. Streuungsmaße sind also geeignete statistische Zahlen, um die Qualität eines Mittelwertes zu beurteilen.

Von den möglichen Streuungsmaßen wird hier die **Spannweite,** die **mittlere lineare Abweichung,** die **mittlere quadratische Abweichung** (Varianz) und die **Standardabweichung** behandelt.

4.1 Spannweite

Die Spannweite R ist ein einfach zu handhabendes Streuungsmaß. Von einer geordneten Datenreihe (Merkmalsausprägung) werden das kleinste und das größte Element herausgezogen. Die Differenz zwischen beiden bestimmt die Spannweite R. Sie ist ein Maß für die Verteilung der Werte um den Mittelwert.

Fortsetzung des Beispiels (vgl. Seite 2):
Vergleich der Spannweite mit den Mittelwerten:

Aus der geordneten Datenreihe von 25 Merkmalsausprägungen werden der kleinste Wert (23) und der größte Wert (59) herausgesucht und die Differenz gebildet; sie gibt die Spannweite R an.

R = 59 – 23 = **36**
Vergleich mit Mittelwerten:
$\bar{x}$ = 38,24 Jahre (vgl. Seite 7)

$\bar{x}$ = 37,8 Jahre (vgl. Seite 8)

Z = 38 Jahre (vgl. Seite 9)

Bei einem arithmetischen Mittel von 38,24 Jahren, bzw. bei einem gewogenen arithmetischen Mittel von 37,8 Jahren und einem ähnlichen Zentralwert von 38 Jahren ist die Spannweite von 36 erheblich. Sie reagiert empfindlich auf Extremwerte und sollte in der Statistik nur in Verbindung mit einer Vergleichsreihe genutzt werden.

Allgemein wird die Spannweite ermittelt:

$\mathbf{R = x_{max} - x_{min}}$

4.2 Mittlere lineare Abweichung

Bei der mittleren linearen Abweichung – auch mittlere arithmetische Abweichung genannt - werden die Differenzen zwischen den einzelnen Merkmalsausprägungen und dem arithmetischen Mittel oder dem Zentralwert **absolut** gesetzt, addiert und die sich ergebende Summe durch die Anzahl der Merkmalsausprägungen dividiert. Die mittlere lineare Abweichung vom arithmetischen Mittel wird symbolhaft durch $\mathbf{d_{\bar{x}}}$ ausgedrückt. Im Gegensatz zur Spannweite wird hier also der Mittelwert in der Berechnung berücksichtigt.

Fortsetzung des Beispiels:

Für die Merkmalsausprägungen (Alter der Arbeitnehmer) aus dem Beispiel von Seite 2 und für das arithmetische Mittel von 38,24 Jahren (vgl. Seite 7) soll die mittlere lineare Abweichung bestimmt werden. Die Summe der absoluten Differenzen wird durch die Anzahl der Merkmalsausprägungen dividiert (25). Die entsprechenden Angaben finden sich in der folgenden Tabelle.

Die absoluten Differenzen aus den Altersangaben $\mathbf{x_i}$ und dem Mittelwert $\bar{x}$ führen dazu, dass sich die sonst auftretenden positiven und negativen Differenzen nicht gegenseitig aufheben.

Tabelle: Alter der Arbeitnehmer zum 01.01.01

laufende Nummer n_i	**Alter der Arbeitnehmer in Jahren x_i**	**Absolute Differenzen $\lvert x_i - \bar{x} \rvert$** zum Mittelwert $\bar{x}$ = 38,24	**Ergebnisse der absoluten Differenzen**
1	23	|23 – 38,24|	15,24
2	23	|23 – 38,24|	15,24
3	23	|23 – 38,24|	15,24
4	23	|23 – 38,24|	15,24
5	25	|25 – 38,24|	13,24
6	25	|25 – 38,24|	13,24
7	25	|25 – 38,24|	13,24
8	29	|29 – 38,24|	9,24
9	35	|35 – 38,24|	3,24

10	35	35 – 38,24	3,24
11	35	35 – 38,24	3,24
12	38	38 – 38,24	0,24
13	38	38 – 38,24	0,24
14	38	38 – 38,24	0,24
15	42	42 – 38,24	3,76
16	42	42 – 38,24	3,76
17	42	42 – 38,24	3,76
18	42	42 – 38,24	3,76
19	47	47 – 38,24	8,76
20	48	48 – 38,24	9,76
21	48	48 – 38,24	9,76
22	56	56 – 38,24	17,76
23	56	56 – 38,24	17,76
24	59	59 – 38,24	20,76
25	59	59 – 38,24	20,76
			240,72

$d_{\bar{x}}$ = 240,72 : 25 = **9,63**

Mit $d_{\bar{x}}$ = 9,63 ist ein Streuungsmaß gefunden, das wesentlich besser als die Spannweite die Streuung der einzelnen Daten um den Mittelwert interpretiert.

Allgemeine wird die mittlere lineare Abweichung durch folgende Gleichung ausgedrückt:

$$d_{\bar{x}} = 1/n \cdot \sum_{i=1}^{n} |x_i - \bar{x}|$$

Fortsetzung des Beispiels:

Die gleiche Berechnung wird für den Zentralwert **Z = 38** (vgl. Seite 9) durchgeführt. Es ergeben sich die in der folgenden Tabelle aufgeführten Werte. Die mittlere lineare Abweichung vom Zentralwert Z (auch Median genannt) wird symbolhaft durch $\mathbf{d_{Med}}$ ausgedrückt.

Tabelle: Alter der Arbeitnehmer zum 01.01.01

laufende Nummer n_i	Alter der Arbeitnehmer in Jahren x_i	Absolute Differenzen $\lvert x_i - Z \rvert$ zum Median Z = 38	Ergebnisse der absoluten Differenzen
1	23	23 – 38	15
2	23	23 – 38	15
3	23	23 – 38	15
4	23	23 – 38	15
5	25	25 – 38	13
6	25	25 – 38	13
7	25	25 – 38	13
8	29	29 – 38	9
9	35	35 – 38	3
10	35	35 – 38	3
11	35	35 – 38	3
12	38	38 – 38	0
13	38	38 – 38	0
14	38	38 – 38	0
15	42	42 – 38	4
16	42	42 – 38	4
17	42	42 – 38	4
18	42	42 – 38	4
19	47	47 – 38	9
20	48	48 – 38	10
21	48	48 – 38	10
22	56	56 – 38	18

23	56	56 – 38	18
24	59	59 – 38	21
25	59	59 – 38	21
			240

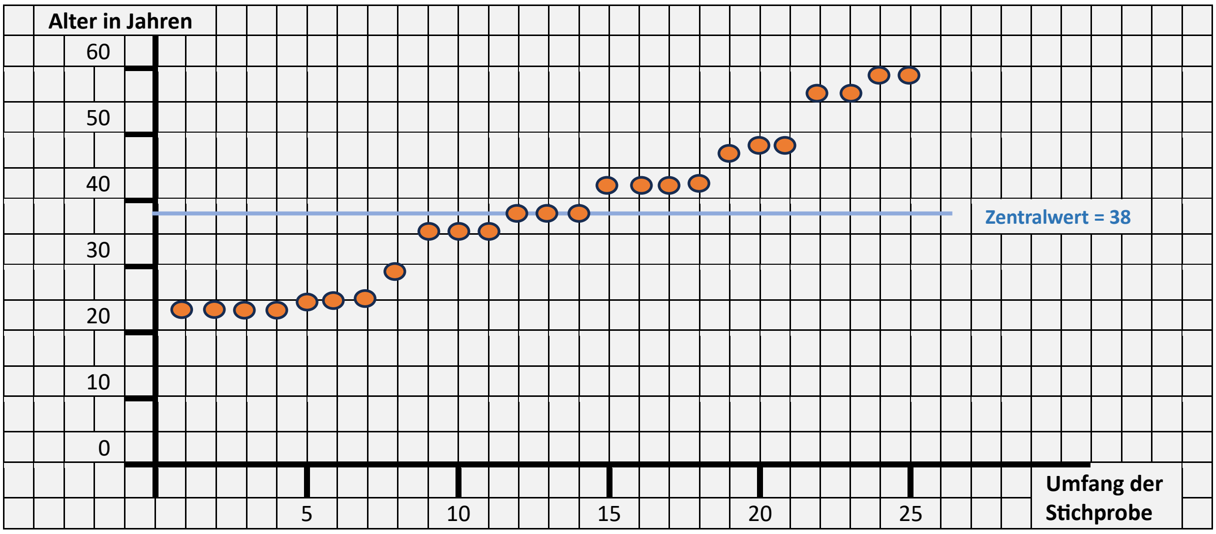
d_{Med} = 240 : 25 = **9,6**

Gegenüber der mittleren linearen Abweichung bezogen auf das arithmetische Mittel (**$d_{\bar{x}}$**) ist der Unterschied zur mittleren linearen Abweichung bezogen auf den Zentralwert (**d_{Med}**) nur unerheblich. Es lässt sich zeigen, dass immer gilt: **$d_{Med} \leq d_{\bar{x}}$.**

Allgemein wird die mittlere lineare Abweichung vom Zentralwert wie folgt ausgedrückt:

$$d_{Med} = 1/n \cdot \sum_{i=1}^{n} |x_i - Z|$$

Die mittlere lineare Abweichung vom Zentralwert lässt sich grafisch wie folgt darstellen:



Die mittlere lineare Abweichung (9,6) vom Zentralwert (38) gibt die Streuung der Altersdaten in angemessener Weise wieder. Je größer die Zahl ist, umso deutlicher werden die Unterschiede hervorgehoben.

Fortsetzung des Beispiels:
Die mittlere lineare Abweichung lässt sich auch auf Häufigkeitstabellen anwenden (vgl. Seite 3). Im folgenden Beispiel wird die mittlere lineare Abweichung vom Zentral Z = 38 anhand der Häufigkeitstabelle dargestellt.

Tabelle: Alter der Arbeitnehmer zum 01.01.01 nach Häufigkeiten

Merkmals-ausprägungen x_i	**Absolute Häufigkeit** Hn_i	**Absolute Differenzen** zum Zentralwert Z = 38 $\lvert x_i - Z\rvert$	$\lvert \mathbf{x_i - Z}\rvert \cdot \mathbf{Hn_i}$
23	4	23 - 38 = 15	60
25	3	25 - 38 = 13	39
29	1	29 - 38 = 9	9
35	3	35 - 38 = 3	9
38	3	38 - 38 = 0	0
42	4	42 - 38 = 4	16
47	1	47 - 38 = 9	9
48	2	48 - 38 = 10	20
56	2	56 - 38 = 18	36

59	2	$\lvert 59 - 38 \rvert = 21$	42
--	25		240

d_{Med} = 240 **:** 25 = **9,6**

Es überrascht nicht, dass in diesem Beispiel das gleiche Ergebnis erreicht wird, wie im Beispiel mit allen Merkmalsausprägungen (vgl. Seite 12). Die Häufigkeitstabelle wird wegen ihrer Übersichtlichkeit und Kürze der ursprünglichen Tabelle vorgezogen.

4.3 Varianz

Die Varianz – auch mittlere quadratische Abweichung genannt – ist in der Statistik ein häufig verwendetes Streuungsmaß. Sie baut auf der mittleren arithmetischen Abweichung auf (vgl. Seite 11), indem sie die absoluten Differenzen zwischen den Merkmalsausprägungen und dem arithmetischen Mittel quadriert, die Ergebnisse zur Summe aufaddiert und diese Summe durch die Anzahl der Merkmalsausprägungen $\mathbf{n_i}$ dividiert. Sie wird mit dem Symbol $\mathbf{s^2}$ gekennzeichnet.

Fortsetzung des Beispiels:

Das Beispiel von Seite 11 wird zur Varianz erweitert.

Tabelle: Alter der Arbeitnehmer zum 01.01.01

laufende Nummer n_i	**Alter der Arbeitnehmer in Jahren** x_i	**Absolute Differenzen** $\lvert x_i - \bar{X} \rvert$ zum Mittelwert $\bar{X}$ = 38,24	**Quadrat der absoluten Differenzen** $\lvert x_i - \bar{X} \rvert^2$	**Ausrechnung und Summenbildung der Quadrate** $\sum (x_i - \bar{X})^2$
1	23	$\lvert 23 - 38{,}24 \rvert$	$\lvert 23 - 38{,}24 \rvert^2$	$15{,}24^2$ = 232,26
2	23	$\lvert 23 - 38{,}24 \rvert$	$\lvert 23 - 38{,}24 \rvert^2$	$15{,}24^2$ = 232,26
3	23	$\lvert 23 - 38{,}24 \rvert$	$\lvert 23 - 38{,}24 \rvert^2$	$15{,}24^2$ = 232,26
4	23	$\lvert 23 - 38{,}24 \rvert$	$\lvert 23 - 38{,}24 \rvert^2$	$15{,}24^2$ = 232,26
5	25	$\lvert 25 - 38{,}24 \rvert$	$\lvert 25 - 38{,}24 \rvert^2$	$13{,}24^2$ = 175,30
6	25	$\lvert 25 - 38{,}24 \rvert$	$\lvert 25 - 38{,}24 \rvert^2$	$13{,}24^2$ = 175,30
7	25	$\lvert 25 - 38{,}24 \rvert$	$\lvert 25 - 38{,}24 \rvert^2$	$13{,}24^2$ = 175,30
8	29	$\lvert 29 - 38{,}24 \rvert$	$\lvert 29 - 38{,}24 \rvert^2$	$9{,}24^2$ = 85,38
9	35	$\lvert 35 - 38{,}24 \rvert$	$\lvert 35 - 38{,}24 \rvert^2$	$3{,}24^2$ = 10,50
10	35	$\lvert 35 - 38{,}24 \rvert$	$\lvert 35 - 38{,}24 \rvert^2$	$3{,}24^2$ = 10,50
11	35	$\lvert 35 - 38{,}24 \rvert$	$\lvert 35 - 38{,}24 \rvert^2$	$3{,}24^2$ = 10,50
12	38	$\lvert 38 - 38{,}24 \rvert$	$\lvert 38 - 38{,}24 \rvert^2$	$0{,}24^2$ = 0,06
13	38	$\lvert 38 - 38{,}24 \rvert$	$\lvert 38 - 38{,}24 \rvert^2$	$0{,}24^2$ = 0,06
14	38	$\lvert 38 - 38{,}24 \rvert$	$\lvert 38 - 38{,}24 \rvert^2$	$0{,}24^2$ = 0,06
15	42	$\lvert 42 - 38{,}24 \rvert$	$\lvert 42 - 38{,}24 \rvert^2$	$3{,}76^2$ = 14,14
16	42	$\lvert 42 - 38{,}24 \rvert$	$\lvert 42 - 38{,}24 \rvert^2$	$3{,}76^2$ = 14,14
17	42	$\lvert 42 - 38{,}24 \rvert$	$\lvert 42 - 38{,}24 \rvert^2$	$3{,}76^2$ = 14,14
18	42	$\lvert 42 - 38{,}24 \rvert$	$\lvert 42 - 38{,}24 \rvert^2$	$3{,}76^2$ = 14,14
19	47	$\lvert 47 - 38{,}24 \rvert$	$\lvert 47 - 38{,}24 \rvert^2$	$8{,}76^2$ = 76,74
20	48	$\lvert 48 - 38{,}24 \rvert$	$\lvert 48 - 38{,}24 \rvert^2$	$9{,}76^2$ = 95,26
21	48	$\lvert 48 - 38{,}24 \rvert$	$\lvert 48 - 38{,}24 \rvert^2$	$9{,}76^2$ = 95,26
22	56	$\lvert 56 - 38{,}24 \rvert$	$\lvert 56 - 38{,}24 \rvert^2$	$17{,}76^2$ = 315,42
23	56	$\lvert 56 - 38{,}24 \rvert$	$\lvert 56 - 38{,}24 \rvert^2$	$17{,}76^2$ = 315,42
24	59	$\lvert 59 - 38{,}24 \rvert$	$\lvert 59 - 38{,}24 \rvert^2$	$20{,}76^2$ = 430,98
25	59	$\lvert 59 - 38{,}24 \rvert$	$\lvert 59 - 38{,}24 \rvert^2$	$20{,}76^2$ = 430,98
				3 388,62

$\mathbf{s^2}$ = 3 388,62 **:** 25 = **135,54**

Die **Varianz** beträgt im Beispiel 135,54. Dies ist ein hoher Wert. Die Varianz macht die Streuungsunterschiede deutlich.

Allgemein lässt sich die **Varianz** wie folgt berechnen:

$$s^2 = 1/n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

4.4 Standardabweichung

Vielfach wird zusätzlich zur Varianz die **Standardabweichung s** (oder **σ**) angegeben, die in der Statistik das bedeutendste Streuungsmaß ist, da sie die größte Aussagekraft besitzt. Die Standardabweichung errechnet sich, indem die Quadratwurzel aus der Varianz gezogen wird:

Standardabweichung s = √135,54 = 11,64

Allgemeine wird die Standardabweichung definiert als

$$s = \sqrt{1/n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Gegenüber der mittleren arithmetischen Abweichung von 9,63 (vgl. Seite 11) hebt die Standardabweichung die Unterschiede zwischen den Merkmalsausprägungen deutlicher hervor.

Die Varianz und die Standardabweichung können als Streuungsmaß auch für den Zentralwert und/oder für Häufigkeitstabellen berechnet werden.

Fortsetzung des Beispiels:

Im folgenden Beispiel soll die Häufigkeitsverteilung der Merkmalsausprägungen von Seite 3 auf die Streuungsmaße Varianz und Standardabweichung untersucht werden; mit $\bar{x}$ = 38,24.

Merkmals-ausprägungen (x_i)	Absolute Häufigkeit n_i	$\lvert x_i - \bar{x} \rvert$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
23	4	|23 - 38,24|	232,26	232,26 · 4 = 929,04
25	3	|25 - 38,24|	175,30	175,30 · 3 = 525,90
29	1	|29 - 38,24|	85,38	85,38
35	3	|35 - 38,24|	10,50	10,50 · 3 = 31,50
38	3	|38 - 38,24|	0,06	0,06 · 3 = 0,18
42	4	|42 - 38,24|	14,14	14,14 · 4 = 56,56
47	1	|47 - 38,24|	76,74	76,74
48	2	|48 - 38,24|	95,26	95,26 · 2 = 190,52
56	2	|56 - 38,24|	315,42	315,42 · 2 = 630,84
59	2	|59 - 38,24|	430,98	430,98 · 2 = 861,96
--	**25**			**3 388,62**

Varianz s^2 = 3 388,62 : 25 = 135,54

Standardabweichung s = √135,54 = 11,64

Es verwundert nicht, dass Varianz und Standardabweichung gegenüber den Ergebnissen aus den Urdaten übereinstimmen. Es lässt sich auch hier vereinfacht mit der Häufigkeitstabelle anstelle der ursprünglichen Tabelle arbeiten.

5 Preisindizes

Indexzahlen finden in volkswirtschaftlichen Statistiken sehr häufig Anwendung, so z. B. als Index der Lebenshaltungskosten, als Index der industriellen Erzeugerpreise, als Index der Großhandelspreisepreise, als Index der Wertpapierkurse u. a.

Preisindizes sind statistische Werte, die die Veränderung einer Größe (Preise eines bestimmten Werkstoffes oder eines bestimmten Erzeugnisses, Umsätze) über mehrere Jahre oder Monate bezüglich einer Basis (Basisjahr, Basismonat) angeben. Sie werden durch Einbeziehung einer Gewichtung (Mengen oder Prozentzahlen) zu gewogenen arithmetischen Mittelwerten.

Beispiel 1: Einfacher Index

In der Thomas Berg KG sind die Gesamtumsätze der letzten drei Jahre in folgender Tabelle zusammengefasst:

Jahre	Gesamtumsatz	Indexzahlen (1. Jahr als Basisjahr)
1.	10.425.000,00 €	100 %
2.	10.633.500,00 €	102 %
3.	10.320.000,00 €	99 %

Nimmt man den Umsatz des 1. Jahres als Basis an und setzt diesen Umsatz gleich 100 %, so ergibt sich für das 2. Jahr eine Umsatzsteigerung um 2 % auf 102 % und für das 3. Jahr eine Umsatzverringerung um 1 % auf 99 %.

Umsatz des 2. Jahres = 10.633.500,00 € **:** 10.425.000,00 € = 1,02 = 102 %

Umsatz des 3. Jahres = 10.320.000,00 € **:** 10.425.000,00 € ≈ 0,99 = 99 %

Indexzahlen erleichtern die Interpretation: Ein Index größer als 100 % bedeutet eine Umsatzsteigerung gegenüber dem Basisjahr, ein Index kleiner als 100 % einen Umsatzrückgang. Zudem gibt der Index eine Vergleichszahl für die Umsatzsteigerung oder den Umsatzrückgang an. So lässt der Index 99 % im 3. Jahr auf einen Umsatzrückgang um 1 % innerhalb des Vergleichszeitraums schließen.

Beispiel 2:

Einen tieferen Einblick in die Umsatzentwicklung gewinnt Herr Berg, wenn er die Umsätze nach Erzeugnisgruppen untersucht. Dazu wählt er das 1. Jahr (Basisjahr) und das 3. Jahr aus und erstellt daraus für die Gehäusetypen GI, GII und GIII aus vorhandenen Zahlen die folgende Umsatztabelle:

Erzeugnis-gruppe	Umsatz des 1. Jahres (Basisjahr)	Umsatz des 3. Jahres	Umsatz-veränderung	Index
Gehäuse GI	4.691.250,00 €	5.160.000,00 €	+ 468.750,00 €	110 %
Gehäuse GII	3.612.000,00 €	3.096.000,00 €	- 516.000,00 €	88 %
Gehäuse GIII	2.121.750,00 €	2.064.000,00 €	- 57.750,00 €	97 %
Gesamtumsatz	10.425.000,00 €	10.320.000,00 €	-105.000,00 €	99 %

Index Gehäuse GI = 5.160.000,00 € **:** 4.691.250,00 € ≈ **1,10 = 110 %**

Index Gehäuse GII = 3.096.000,00 € **:** 3.612.000,00 € ≈ **0,88 = 88 %**

Index Gehäuse GIII = 2.064.000,00 € **:** 2.121.750,00 € ≈ **0,97 = 97 %**

Die obige Darstellung macht die Umsatzentwicklung des 3. Jahres bezogen auf das Basisjahr für einzelne Erzeugnisgruppen deutlich. Es ist nicht nur abzulesen, wie sich die Umsätze in Euro für die einzelnen Erzeugnisgruppen entwickelt haben, sondern auch in Indexzahlen. Danach gibt der Index von 99 % für den Gesamtumsatz nur eine grobe Richtschnur an. Über die Aufgliederung in einzelne Gruppenumsätze wird ein gänzlich anderes Bild vermitteln: Die Gehäusegruppe GI ist nicht nur die stärkste Umsatzgruppe, sondern hat auch einen deutlichen Zuwachs auf 110 % - also um 10 % - verzeichnet, der nicht durch die rückläufigen Umsätze der übrigen Gruppen ausgeglichen wird. Besonders auffallend ist der Umsatzrückgang beim Gehäuse GII auf nunmehr 88 %.

Beispiel 3: Erweiterter Index

Herr Berg will zusätzlich die Umsatzstärke der einzelnen Erzeugnisgruppen am Gesamtumsatz und die Veränderung der Verkaufspreise in seine Indexberechnung einbeziehen. Dazu wählt er das erste Jahr als Basisjahr aus und errechnet zunächst die Umsatzstärke:

Umsatzanteil GI = 4.691.250,00 € **:** 10.425.000,00 € = **0,45 = 45,0 %**

Umsatzanteil GII = 3.612.000,00 € **:** 10.425.000,00 € = **0,346 = 34,6 %**

Umsatzanteil GIII = 2.121.750,00 € **:** 10.425.000,00 € = **0,204 = 20,4 %**

Mit dieser Gewichtung ist es ihm möglich, unter Einbeziehung der Entwicklung der **Verkaufspreise** einen aussagefähigen **Preisindex** für den Gesamtumsatz des 3. Jahres bezogen auf das Basisjahr zu erhalten. In der folgenden Tabelle sind neben den Umsatzzahlen auch die durchschnittlichen Verkaufspreise je Gehäuse für das 1. Jahr und das 3. Jahr verzeichnet:

Erzeugnisgruppe	1. **Jahr** (Basisjahr)		**3. Jahr**	
	Verkaufspreise	**Umsätze**	**Verkaufspreise**	**Umsätze**
Gehäuse GI	46,91 €	4.691.250,00 €	46,91 €	5.160.000,00 €
Gehäuse GII	45,15 €	3.612.000,00 €	40,00 €	3.096.000,00 €
Gehäuse G III	28,29 €	2.121.750,00 €	27,52 €	2.064.000,00 €
Gesamtumsatz		10.425.000,00 €		10.320,000,00 €

Die Veränderung der Verkaufspreise wird berechnet, indem man den Preis des 3. Jahres durch den Preis des 1. Jahres dividiert:

Preisänderung GI = 46,91 € **:** 46,91 € = **1,0**

Preisänderung GII = 40,00 € **:** 45,15 € = **0,886**

Preisänderung GIII = 27,52 € **:** 28,29 € = **0,973**

Im letzten Schritt wird die Gewichtung (Umsatzanteile) mit den Preisänderungen verbunden, indem für jede Erzeugnisgruppe das Produkt aus Umsatzanteil und Preisänderung gebildet und die so gebildeten Produkte additiv zu einem einheitlichen Preisindex zusammengefasst wird. Es ergibt sich folgender **Preisindex für den Gesamtumsatz**:

Preisindex = $0,45 \cdot 1 + 0,346 \cdot 0,886 + 0,204 \cdot 0,973$ =

Preisindex = 0,45 + 0,3066 + 0,1985

Preisindex = 0,955

Gegenüber dem in Beispiel 1 ermittelten Index von 0,99 (99 %) ist der Preisindex mit 0,955 (95,5 %) wesentlich genauer, da er sowohl die Preisänderungen als auch die Gewichtung der einzelnen Erzeugnisgruppen berücksichtigt.