

DAA Wirtschafts-Lexikon

Betriebsstatistik

A. Darstellung des statistischen Materials¹

1. Die Tabelle

Eine Tabelle muss so gestaltet sein, dass sie die für die Fragestellung relevanten Informationen übersichtlich wiedergibt. Sie muss klar und eindeutig aufgebaut sein und alle Informationen enthalten, die für das Verständnis notwendig sind.

Die wichtigsten *Anforderungen* an eine Tabelle sind:

- Die Überschrift muss den Geltungsbereich der Daten eindeutig definieren.
- Die Tabellenummer ist bei größeren Zusammenhängen wichtig, damit im Text auf die entsprechende Tabelle Bezug genommen werden kann.
- Der Tabellenkopf gibt den Inhalt der Spalten an und beschreibt das Merkmal bzw. die Merkmalsausprägung, deren Häufigkeiten in der zugehörigen Spalte aufgeführt werden.
- Die Vorspalte enthält die entsprechenden Informationen über die einzelnen Zeilen.
- Die Maßeinheiten, in der die Daten der Tabelle gemessen werden, müssen in der Tabelle unbedingt angegeben werden.
- Die Quellen, aus denen die Daten stammen, sind in der Fußnote aufzuführen.
- Bei großen Tabellen kann eine Nummerierung der Spalten und Zeilen sinnvoll sein, damit bei der Beschreibung und Interpretation der Tabelle einfacher auf die wesentlichen Ergebnisse hingewiesen werden kann.

Eine Tabelle hat gegenüber den graphischen Darstellungen den Vorteil, dass die Zahlenwerte exakt abgelesen werden können, allerdings wird sie desto *unübersichtlicher* je mehr Merkmale sie enthält. auch bei hohen Zahlenwerten oder gebrochenen Zahlen ist das Verständnis erschwert.

2. Das Säulendiagramm

Schaubilder müssen - wie Tabellen - eindeutig und *übersichtlich* sein. Da sie die Aufgabe haben, auf einen Blick zu informieren, sollten sie nicht mit Informationen überladen werden.

¹ Quellen: HOLLAND, H./SCHARNBACHER, K.: Grundlagen der Statistik. Gabler Verlag, Ludwigshafen.
SCHWARZE, J.: Grundlagen der Statistik, NWB Verlag, Herne 2009.

Ein Säulendiagramm besteht aus einem System mit zwei Achsen, wobei die **Abszisse** (die waagerechte Achse) das dazustellende Merkmal mit den entsprechenden Merkmalsausprägungen enthält. An der *Ordinate* werden die Häufigkeiten abgetragen. Über jeder Merkmalsausprägung wird parallel zur Ordinate eine Linie gezogen, deren Höhe der Häufigkeit der Merkmalsausprägung entspricht.

Das Säulendiagramm ist eine *höhenproportionale* Darstellung, da allein die Höhe oder Länge des Stabes die Information über die darzustellende Häufigkeit enthält. Auch nominalskalierte Merkmale, deren Ausprägungen sich nicht in eine Reihenfolge bringen lassen, können in ein Stabdiagramm eingezeichnet werden.

Säulendiagramme sind höhenproportionale Darstellungen, die schon bei nominalskalierten Daten verwendet werden können.

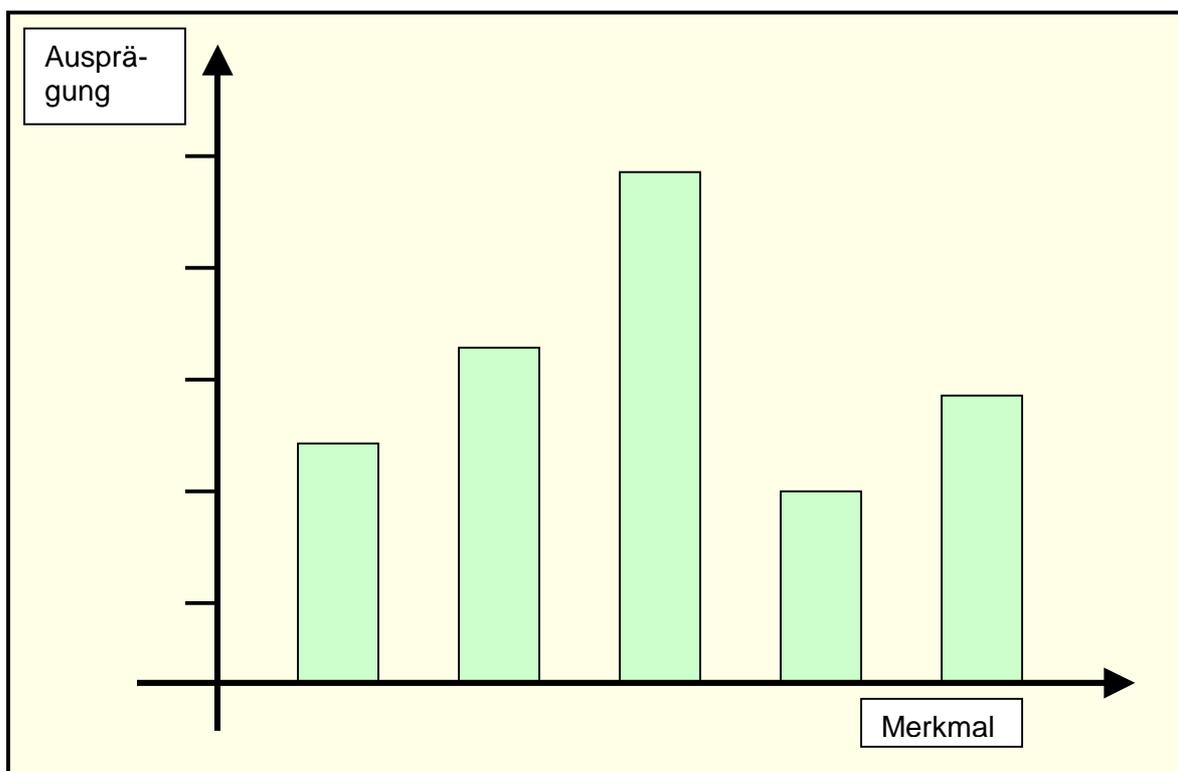


Abb. 1: Säulendiagramm

3. Das Histogramm

Auch bei einem Histogramm werden die Daten in einem Koordinatensystem eingetragen, das auf der Abszisse die Merkmalsausprägungen und auf der Ordinate (y-Achse) die Häufigkeiten darstellt. Im Unterschied zum Stabdiagramm wird die Häufigkeit nicht durch eine Höhe, sondern durch eine *Fläche* wiedergegeben.

Vor allem für *klassifizierte* Daten ist das Histogramm gut geeignet. Es bildet die Häufigkeiten bei klassifizierten Daten durch Rechtecke ab, deren Breite der Klassenbreite entspricht. Die Höhe ist bei Daten mit gleichen Klassenbreiten mit der Häufigkeit identisch, denn bei gleicher Grundlinie ist die Höhe eines Rechtecks proportional zur Fläche. Falls *ungleiche Klassenbreiten* vorliegen, muss die Höhe der Säulen in Abhängigkeit von der Breite berechnet werden.

Histogramme sind flächenproportionale Darstellungen, die sich vor allem bei klassifizierten Daten eignen.

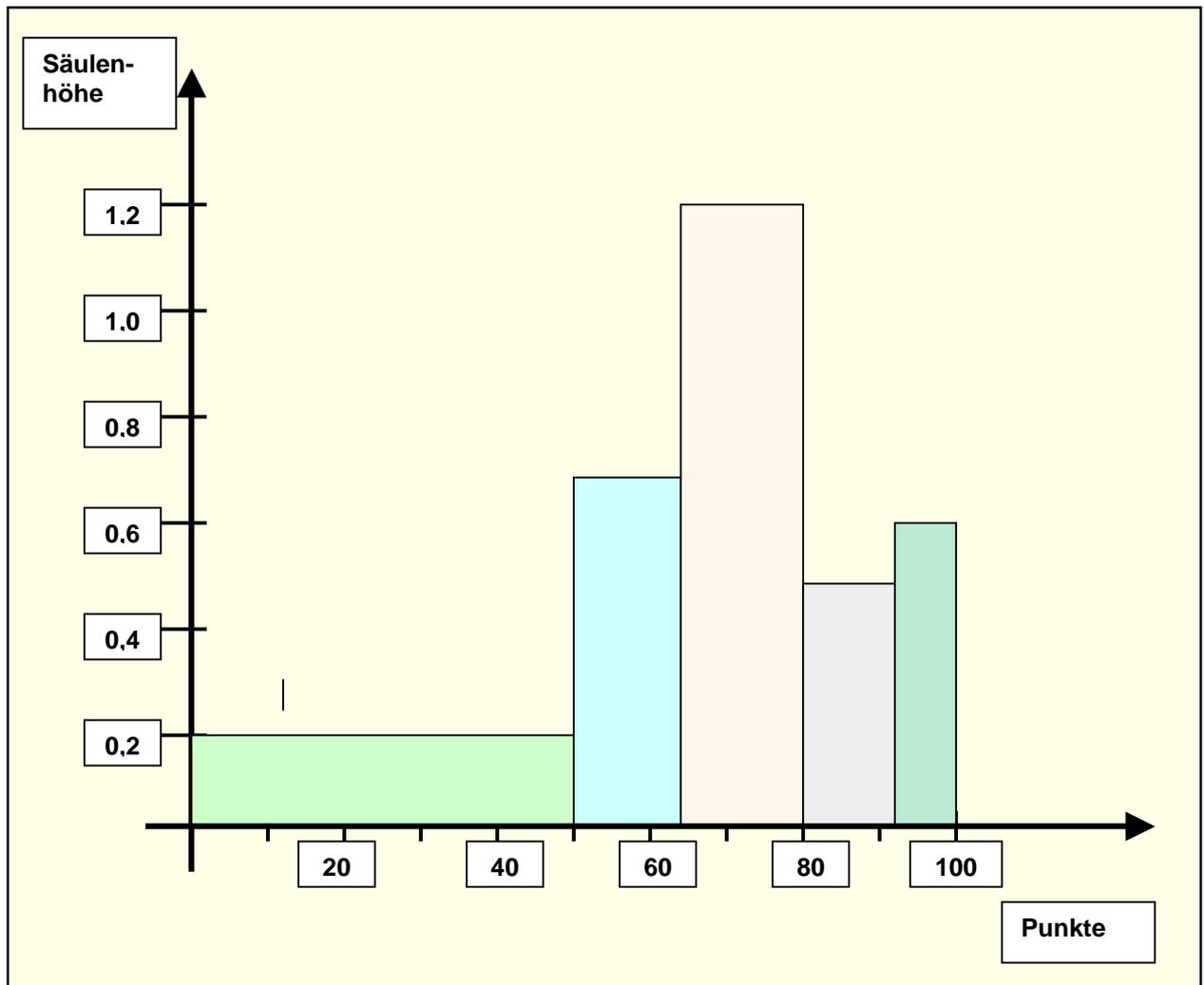


Abb. 2: Histogramm der Klausurergebnisse (Beispiel)

4. Das Kreisdiagramm

Ein weiteres in den gedruckten Medien häufig zu findendes Diagramm, da auch bei *nominalskalierten* Merkmalen verwendet werden kann, ist das Kreisdiagramm.

Ein Kreisdiagramm ist eine flächenproportionale Darstellung, bei der die Kreisfläche der Grundgesamtheit entspricht. Der Kreis wird in einzelne

Sektoren eingeteilt, die der Häufigkeit der jeweiligen Merkmalsausprägung entsprechen.

Der Kreisumfang, der aus 360 Winkelgraden besteht, stellt die Größe der Grundgesamtheit dar. Die einzelnen Merkmalsausprägungen werden so in das Kreisdiagramm eingetragen, dass deren Anteile am Kreisumfang den prozentualen Anteilen der Merkmalsausprägungen an der Grundgesamtheit entsprechen.

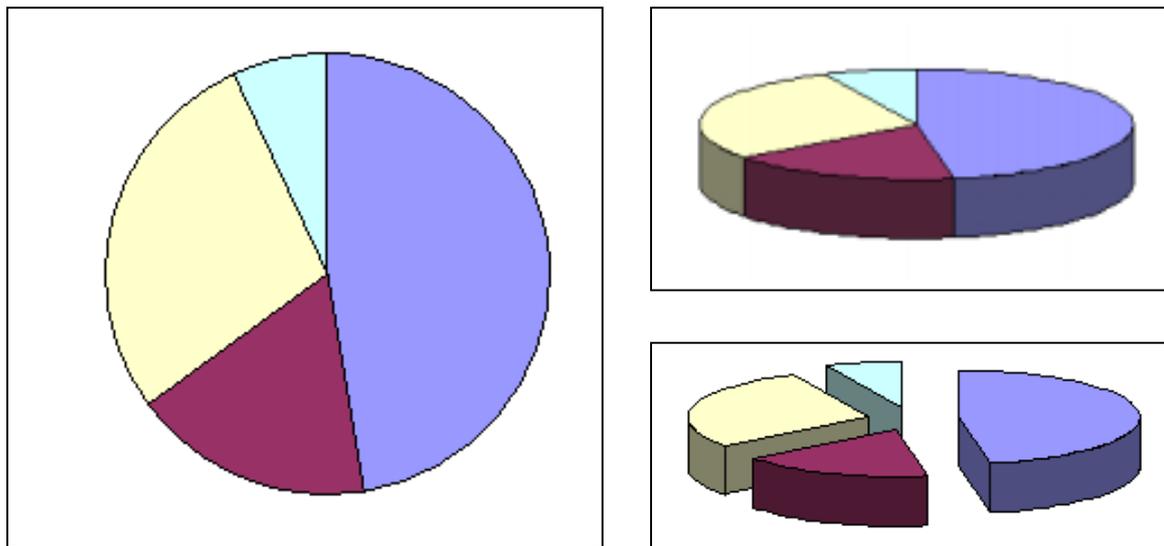


Abb.3 : Kreisdiagramme

5. Das Polygon

Wenn die Anzahl der Klassen in die die Daten eingeordnet sind, sehr **groß** ist, ist das Polygon zur Darstellung der Verteilung besser als das Histogramm geeignet. Das Polygon kann direkt aus dem Histogramm entwickelt werden.

Bei klassifizierten Daten ist die Verteilung der Werte *innerhalb* einer Klasse nicht mehr bekannt. In diesem Fall trifft man die Annahme, dass die Werte in jeder Klasse gleichmäßig über die gesamte Klassenbreite verteilt sind und somit der *Mittelpunkt* als repräsentativ für die Klasse gelten kann.

Man kann dann über die Klassenmitte auf der Abszisse die Häufigkeit abtragen. Es ergibt sich für jede Klasse ein Punkt im Koordinatensystem. durch die Verbindung der Punkte erhält man das *Polygon* oder den Polygonzug. Das Polygon stellt noch keine mathematische Funktion dar.

Wenn man die Klassenbreiten verkleinert und damit die Anzahl der Klassen erhöht oder bei sehr vielen Beobachtungswerten das Merkmal direkt ohne Klassifizierung darstellt, ergibt sich letztendlich eine stetige *Verteilungskurve*.

Das Polygon ist eine flächenproportionale Darstellung, bei der die Häufigkeiten an der Klassenmitte eingetragen werden.

Aus dem Histogramm lässt sich das Polygon entwickeln, indem man die Mittelpunkte der oberen Säulenbegrenzungen miteinander verbindet.

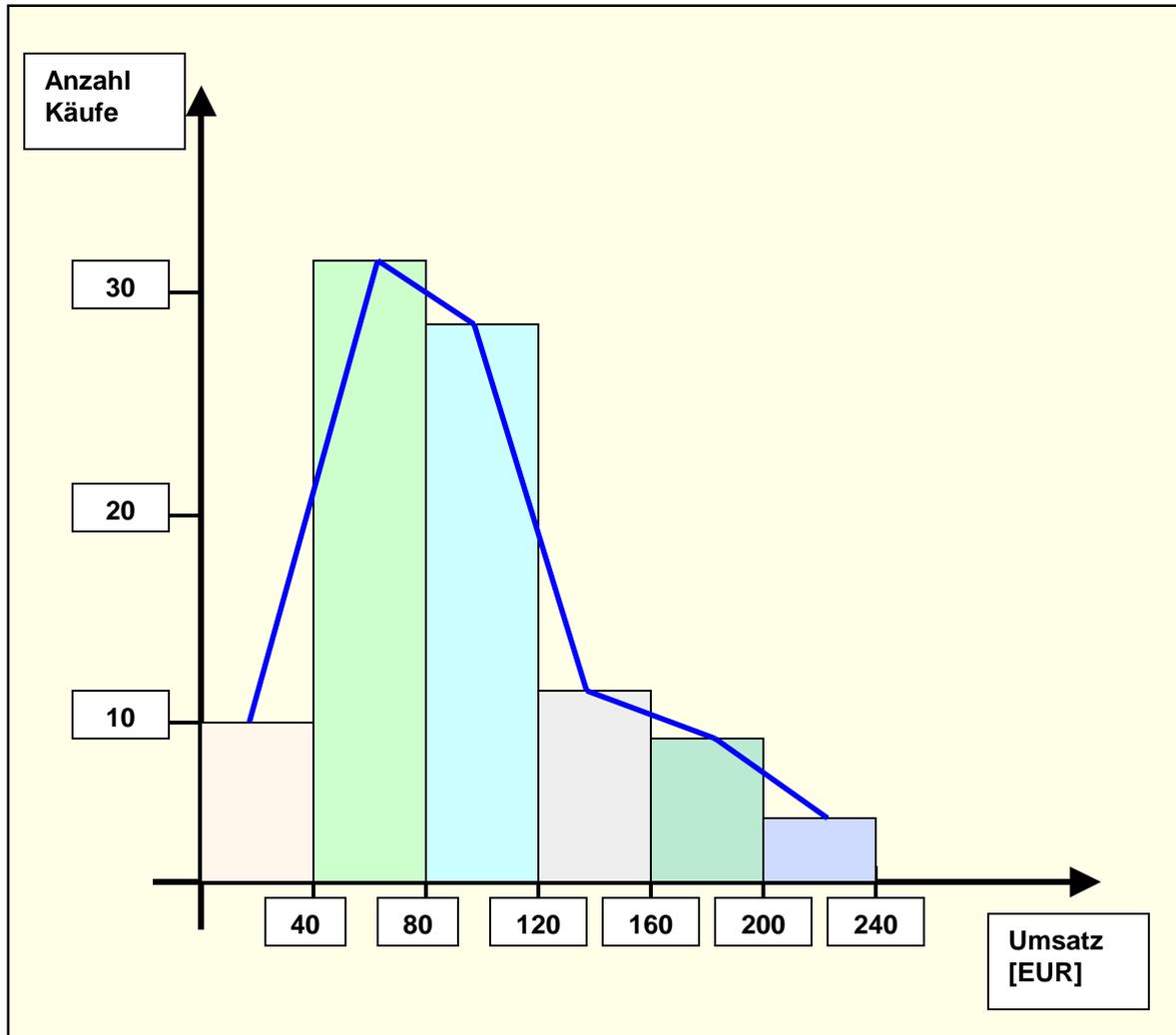


Abb. 4: Polygon

6. Die Summenkurve

Bei einer Häufigkeitsverteilung ist oftmals nicht nur die Frage nach der Häufigkeit in den einzelnen Klassen von Bedeutung, sondern auch die Frage, wie viele der Elemente **über** bzw. **unter** einem bestimmten Wert liegen. Die Summenkurve, die diese Frage beantwortet, lässt sich nicht für nominalskalierte Daten aufstellen, da die Elemente dazu in einer Reihenfolge gebracht werden müssen.

Die Summenkurve wird berechnet, indem die Häufigkeiten der Merkmalsausprägungen in aufsteigender oder abfallender Richtung nacheinander summiert werden.

Zwei Fragestellungen lassen sich beantworten:

1. Wie groß ist die Anzahl der Elemente, die *unter* ("weniger als") einem bestimmten Merkmalswert liegen?

Die Summation beginnt zur Beantwortung dieser Frage in der Klasse mit der kleinsten Merkmalsausprägung, also aufsteigen.

2. Wie groß ist die Anzahl der Elemente, die *über* ("mehr als") einem bestimmten Merkmalswert liegen?

Die Summation beginnt in der Klasse des größten Merkmalswertes und geht dann **abfallend** bis zur ersten Klasse zurück.

Beispiel:

Die Vorgehensweise soll wieder an einem Zahlenbeispiel erläutert werden.

Umsatz in Euro	Anzahl der Käufe	aufsteigend kumuliert	abfallend kumuliert
unter 40	10	10	95
40 bis unter 80	32	42	85
80 bis unter 120	28	70	53
120 bis unter 160	12	82	25
160 bis unter 200	9	91	13
200 bis unter 240	4	95	4
Insgesamt	95	-	-

Zu Frage 1:

Wenn die Frage stellt wird, wie groß die Anzahl der Käufe mit einem Umsatz *unter* 120 EUR ist, muss man die Häufigkeiten der ersten 3 Klassen summieren. Man beginnt in der Klasse mit dem kleinsten Merkmalswert und summiert die vier Häufigkeiten *aufsteigend*, bis der gesuchte Wert erreicht ist. Die Anzahl der Käufe unter 120 EUR beträgt als $10 + 32 + 28 = 70$.

Zu Frage 2:

Zur Beantwortung der Frage, wie groß die Anzahl der Käufe mit einem Umsatz *über* 120 EUR ist, müssen die Häufigkeiten der letzten drei Klassen summiert

werden. Die Summation beginnt in der letzten Klasse (4 Käufe liegen über 200 EUR) und arbeitet sich schrittweise *abfallend* bis zu dem gesuchten Wert vor. Die Anzahl der Käufe über 120 EUR beträgt $4 + 9 + 12 = 25$.

Die graphische Darstellung verdeutlicht das Ergebnis:

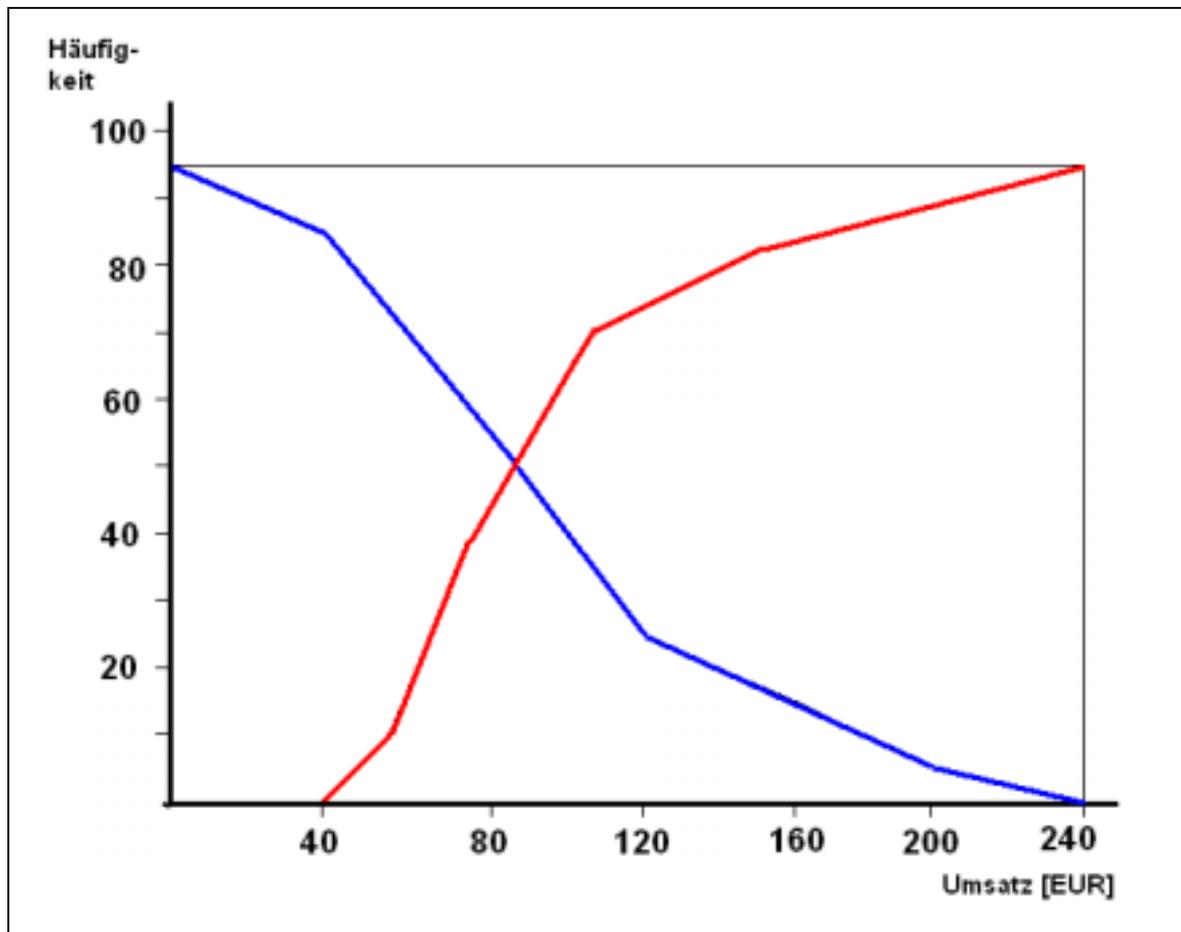


Abb. 5: Summenkurve der Umsatzverteilung

Die Kurve der *aufsteigenden* Kumulation ist vom Ursprung ausgehend zu interpretieren. Beispielsweise lässt sich ablesen, dass 82 Käufe einen Umsatz von unter 160 EUR aufweisen. an der Art der Fragestellung wird deutlich, dass bei einer aufsteigenden Summenkurve die Häufigkeit im Unterschied zu allen anderen besprochenen graphischen Darstellungen auf der **Klassenobergrenze** abzutragen ist.

Bei der *abfallenden* Summenkurve sind dagegen die Häufigkeiten an der Klassenuntergrenze einzureichen. Hier lautet die Interpretation beispielsweise: 13 Käufe haben eine Umsatzhöhe von mehr als 260 EUR.

Die einzelnen Werte werden verbunden, so dass man einen durchgehenden Kurvenzug erhält und auch die Häufigkeiten für Zwischenwerte ablesen kann.

Zum Vergleich verschiedener Grundgesamtheiten empfiehlt es sich, nicht die absoluten sondern die *relativen* prozentualen Häufigkeiten einzuzeichnen.

Beispiel:

Ein Handwerksunternehmen hat im letzten Jahr die in folgender Tabelle aufgeführte Verteilung von Aufträgen festgestellt. Durch eine Summenkurve soll die Tendenz graphisch veranschaulicht werden.

Auftragshöhe in EUR	Anzahl Aufträge	Anteil Aufträge	kumulierte Anteile	
			aufsteigend	abfallend
0 bis unter 1000	40	8 %	8 %	100 %
1000 bis unter 2000	50	10 %	18 %	92 %
2000 bis unter 3000	100	20 %	38 %	82 %
3000 bis unter 4000	160	32 %	70 %	62 %
4000 bis unter 5000	90	18 %	88 %	30 %
5000 bis unter 6000	60	12 %	100 %	12 %
Insgesamt	500	100 %	-	-

Die Kurven müssen sich senkrecht zu 100 % addieren und demnach bei 50 % schneiden. Aus der nachstehenden Abbildung lässt sich ablesen, dass etwa 28 % der Aufträge weniger als 2 500 EUR lauten und etwa 46 % über 3 500 EUR liegen.

7. Die Konzentrationskurve

Die *Konzentrationskurve* oder *Lorenz-Kurve* hat gegenüber der Summenkurve den Vorteil, dass sie die Häufigkeiten *zweier* Massen darstellen kann. In den bisher behandelten Darstellungsformen wurde lediglich die Ordinate zum Abtragen von Häufigkeiten genutzt. Für die Konzentrationskurve unterteilt man ebenso die Abszisse (die x-Achse) und trägt dort die Häufigkeiten des zweiten Merkmals ein.

Da nun keine Achse zur Darstellung des Maßstabs mehr zur Verfügung steht, trägt man auf beiden Achsen *summierte relative* Häufigkeiten ein.

Die Konzentrationskurve erlaubt die Darstellung zweier Merkmale in einem Diagramm. An den Achsen werden die kumulierten relativen Häufigkeiten der beiden zu vergleichenden Merkmale eingetragen.

B. Statistische Maßzahlen

1. Mittelwerte

Statistische **Maßzahlen** haben die Aufgabe, die in den Daten gegebenen Informationen noch stärker zu verdichten und in einigen wenigen Zahlenwerten die Besonderheiten hervorzuheben.

1.1 Der Modus

Der **Modus** gehört zur Gruppe der Mittelwerte, die die Aufgabe haben, die Lage der Verteilung auf der Abszisse anzugeben. Mittelwerte geben an, an welcher Stelle der x-Achse der Schwerpunkt der Verteilung liegt.

Der Modus (M_o) ist derjenige Wert, der in einer Verteilung am häufigsten vorkommt. Er kann bei nicht klassifizierten Werten direkt abgelesen werden. Bei in Klassen eingeteilten Werten muss er durch eine Formel berechnet werden.

Beispiel:

Notenverteilung der Ergebnisse einer schriftlichen Prüfung

Zensuren	1	2	3	4	5	Summe
Anzahl	5	7	14	13	6	45

Die Note 3 kommt am häufigsten vor, damit ist 3 der Modus, $M_o = 3$.

Der Modus hat den Vorteil, dass er direkt ohne Rechenarbeit aus der Tabelle abgelesen werden kann. Ein wichtiger **Nachteil** liegt darin, dass der häufigste Wert nur von den Größenverhältnissen an einer Stelle beeinflusst wird. Damit werden die Informationen, die in den Daten enthalten sind, nur ungenügend ausgeschöpft.

1.2 Der Median

Auch der **Median** (M_z) oder **Zentralwert** gehört zur Gruppe der Mittelwerte. Er kann berechnet werden, wenn die Merkmalswerte der Größe nach geordnet werden können, also zumindest ordinalskaliert sind.

Der Median ist die Merkmalsausprägung desjenigen Wertes, der eine der Größe nach geordnete Reihenfolge halbiert.

Wenn die Grundgesamtheit eine **ungerade** Anzahl von Werten umfasst, so gibt es einen Wert der in der Mitte steht. Es ist der Wert mit der **Ordnungsnummer**

$$z = \frac{n+1}{2}.$$

Beispiel:

Ein Warenhaus "W" hat in seinen 7 Filialen in einem Bundesland im letzten Jahr die folgenden Umsätze (in Mio. EUR) erzielt:

67, 75, 54, 115, 57, 84, 76.

Um den Median zu ermitteln, müssen die Werte zunächst in eine Reihenfolge gebracht werden:

54, 57, 67, 75, 76, 84, 115.

Die Reihe umfasst $n = 7$ Elemente. Die Ordnungsnummer des Median ist

$$z = \frac{7+1}{2} = 4.$$

Das vierte Element ist das in der Mitte stehende. Der Median beträgt damit $M_z = 75$.

1.3 Das arithmetische Mittel

Der am häufigsten berechnete Mittelwert, der im allgemeinen Sprachgebrauch einfach als Mittelwert oder Durchschnitt bezeichnet wird, ist das **arithmetische Mittel**.

Das **arithmetische Mittel** (\bar{x}) entspricht der **Summe der Merkmalsausprägungen dividiert durch deren Anzahl**.

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

Beispiel:

Der Durchschnittsumsatz der 7 Filialen des Warenhauses W ist zu berechnen.

Umsatz der 7 Filialen in Mio. EUR: 54, 57, 67, 75, 76, 84, 115.

$$x = (54 + 57 + 67 + 75 + 76 + 84 + 115) / 7 = 528 / 7 = 75,429 \text{ Mio. EUR.}$$

Der Durchschnittsumsatz beträgt **75,4 Mio. EUR**.

Wenn eine Grundgesamtheit mehrere gleiche Merkmalswerte enthält, so lässt sich die Berechnung vereinfachen, indem man jede Merkmalsausprägung mit der Häufigkeit ihres Auftretens gewichtet und das **gewogene arithmetische Mittel** ermittelt.

Beispiel:

Aus der Altersverteilung einer Gruppe von Auszubildenden soll das Durchschnittsalter berechnet werden:

Alter in Jahren	17	18	19	20
Häufigkeit	10	18	15	7

Bei der Berechnung des gewogenen arithmetischen Mittels werden die

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

mit den Häufigkeiten $h_1, h_2, h_3, \dots, h_n$

gewichtet.

Die Summe der Produkte aus Merkmalsausprägung und Häufigkeit wird durch die Summe der Häufigkeiten dividiert.

$$\bar{x} = \frac{\sum x_i \cdot h_i}{\sum h_i}$$

Für das Beispiel bedeutet das

$$\bar{x} = \frac{17 \cdot 10 + 18 \cdot 18 + 19 \cdot 15 + 20 \cdot 7}{10 + 18 + 15 + 7} = \frac{919}{50} = 18,38 \text{ Jahre}$$

Beispiel:

Bei der Berechnung des Durchschnittsumsatzes für das Beispiel des Warenhauses "W" muss von den Klassenmitten ausgegangen werden, da die Verteilung der Elemente innerhalb der Klassen nicht bekannt ist.

Umsatz in Euro	Klassenmitte	Anzahl der Käufe
unter 40	20	10
40 bis unter 80	60	32
80 bis unter 120	100	28
120 bis unter 160	140	12
160 bis unter 200	180	9
200 bis unter 240	220	4
Insgesamt	-	95

$$\bar{x} = \frac{20 \cdot 10 + 60 \cdot 32 + 100 \cdot 28 + 140 \cdot 12 + 180 \cdot 9 + 220 \cdot 4}{95} = \frac{9100}{95} = 95,79$$

Der Durchschnittsumsatz beträgt **95,79** EUR.

1.4 Das geometrische Mittel

Um Entwicklungstendenzen im *Zeitablauf* zu charakterisieren, benutzt man das geometrische Mittel, bei dem alle Merkmalswerte miteinander multipliziert werden, um anschließend die n-te Wurzel aus diesem Produkt zu ziehen.

Das geometrische Mittel wird zur Analyse von Wachstumsraten im Zeitablauf berechnet.

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

Wenn jährliche Zuwachsraten gegeben sind, kann mit Hilfe des geometrischen Mittels die durchschnittliche jährliche Zuwachsrate ermittelt werden.

Beispiel:

Die Absatzmenge des Warenhauses "W" hat sich in den vergangenen Geschäftsjahren (Gj.) wie folgt entwickelt: Vom Gj. 1 zum Gj. 2: **8 %** Steigerung, vom Gj. 2 zum Gj. 3: **15 %** Steigerung, vom Gj. 3 zum Gj. 4: **10 %** Steigerung, vom Gj. 4 zum Gj.5: **8 %** Steigerung.

Es handelt sich um relative prozentuale Werte, die umgewandelt werden müssen.

vom Gj. 1 zum Gj. 2: **1,08**

vom Gj. 2 zum Gj. 3: **1,15**

vom Gj. 3 zum Gj. 4: **1,10**

vom Gj. 4 zum Gj. 5: **1,05**

Diese Werte sind in die Formel einzusetzen:

$$G = \sqrt[4]{1,08 \cdot 1,15 \cdot 1,10 \cdot 1,05} = \sqrt[4]{1,43451} = 1,0944$$

Die durchschnittliche Zuwachsrate beträgt **9,44 %**.

2. Streuungsmaße

Die Berechnung eines Mittelwertes reicht nicht aus, um eine Verteilung vollständig zu beschreiben. Mittelwerte geben die Lage einer Verteilung auf der Abszisse und damit die mittlere Tendenz an. sie sagen aber nichts über die *Abweichung* der Einzelwerte von diesem Mittelwert.

Die statistischen Kennzahlen, die über die Streuung Auskunft geben, sind die **Streuungsmaße**.

2.1 Spannweite

Die Spannweite ist die Differenz zwischen dem größten und dem kleinsten Merkmalswert, der in einer Verteilung vorkommt.

Bei klassifizierten Daten ist darauf zu achten, dass man nicht von den Klassenmitten, sondern vom kleinsten Wert der ersten Klasse und dem größten Wert der letzten Klasse auszugehen hat.

Beispiel:

Im Falle des Warenhauses „W“ beträgt die Spannweite $240 - 0 = 240$ EUR.

Das Beispiel verdeutlicht, dass die Spannweite nur dazu dienen kann, einen schnellen Überblick über eine Verteilung zu geben. Sie wird von den extremen Werten verfälscht, da nur diese beiden **Extremwerte** in die Berechnung eingehen.

2.2 Die mittlere Abweichung

Während bei der Ermittlung der Spannweite nur die beiden Extremwerte berücksichtigt werden, gehen in die Berechnung der folgenden Streuungsmaße **alle** Beobachtungswerte ein.

Für die **mittlere Abweichung** wird die Differenz zwischen jedem Beobachtungswert und dem arithmetischen Mittel berechnet und aufaddiert. Da sich bei der Addition aber positive und negative Abweichungen aufheben würden, geht man von der Summe der *absoluten* Beträge aus. Aus allen negativen Werten werden positive. Die Summe wird schließlich durch die Anzahl der Werte dividiert, um zu einem Mittelwert der Abweichungen zu kommen.

Die mittlere Abweichung (e) ist das arithmetische Mittel aus den absoluten Beträgen der Abweichungen aller Beobachtungswerte einer Verteilung vom arithmetischen Mittel.

Bei der Berechnung ist, wie beim arithmetischen Mittel, zwischen dem **ungewogenen** und dem **gewogenen** Fall, bei dem mehrere gleiche Merkmalswerte vorkommen, zu unterscheiden.

$$e = \frac{\sum x_i - x}{n} \quad \text{ungewogener Fall}$$

$$e = \frac{\sum x_i - x \cdot h_i}{\sum h_i} \quad \text{gewogener Fall}$$

Die *Summationsgrenzen* lauten üblicherweise **i = 1 bis n** und werden der Einfachheit halber weggelassen. Anstelle des arithmetischen Mittels kann auch von einem beliebigen anderen Mittelwert, wie dem Modus Median, ausgegangen werden.

Beispiel:

Berechnen Sie die mittlere Abweichung des Umsatzes im Warenhaus „W“ vom Durchschnittsumsatz, der mit **95,79** EUR ermittelt wurde.

Umsatz in EUR	Klassenmitte	Anzahl f_j	$x_j - x$	$ x_j - x \cdot f_j$
unter 40	20	10	-75,79	757,90
40 bis unter 80	60	32	-35,79	1145,28
80 bis unter 120	100	28	4,21	117,88
120 bis unter 160	140	12	44,21	530,52
160 bis unter 200	180	9	84,21	757,89
200 bis unter 240	220	4	124,21	496,84
Insgesamt	-	95	-	3806,31

$$e = \frac{3\,806,31}{95} = 40,0664$$

Die mittlere Abweichung der Umsatzhöhe der 96 Käufe vom arithmetischen Mittel beträgt **e = 40,07 EUR**. Je größer der Wert, desto größer die Streuung.

2.3 Varianz und Standardabweichung

Die wichtigsten Streuungsparameter, die auch in der Praxis meist verwendet werden, sind die **Varianz σ^2** und die Quadratwurzel daraus, die **Standardabweichung σ** (Sigma).

Bei der Berechnung geht man ganz ähnlich vor wie bei der mittleren Abweichung. Statt der betragsmäßigen Betrachtung der Differenzen zwischen den Beobachtungswerten und dem arithmetischen Mittel werden bei der Varianz die Differenzen *quadriert*, damit sich positive und negative Abweichungen nicht saldieren.

Die Varianz ist die Summe der Abweichungsquadrate aller Merkmalswerte einer Verteilung von ihrem arithmetischen Mittel, dividiert durch die Anzahl der Merkmalswerte.

Auch hier ist der **gewogene** und **ungewogene** Fall zu unterscheiden.

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad \text{ungewogener Fall}$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot h_i}{\sum h_i} \quad \text{gewogener Fall}$$

Beispiel:

Berechnung der Varianz für den Umsatz des Warenhauses "W":

Umsatz in EUR	Klassen- mitte	Anzahl h_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2 \cdot h_i$
unter 40	20	10	-75,79	57 441,2410
40 bis unter 80	60	32	-35,79	40 989,5712
80 bis unter 120	100	28	4,21	496,2748
120 bis unter 160	140	12	44,21	23 454,2892
160 bis unter 200	180	9	84,21	63 821,9169
200 bis unter 240	220	4	124,21	61 712,4964
Insgesamt	-	95	-	247 915,7895

$$\sigma^2 = \frac{247\,915,7895}{95} = 2\,609,6399$$

Die Varianz beträgt 2 609,64. Bei der *Interpretation* des Ergebnisses treten Schwierigkeiten auf, da die Varianz durch die Quadratur bei der Berechnung als *Dimension* nun das Quadrat der ursprünglichen Dimension hat. Im vorliegenden Beispiel beträgt die Dimension also EUR².

Um diese Interpretationsprobleme zu umgehen und um das eigentliche Streuungsmaß zu erhalten, wird die *Standardabweichung* als Quadratwurzel aus der Varianz berechnet.

$$\sigma = \sqrt{\sigma^2}$$

Für das obige Beispiel beträgt die Standardabweichung

$$\sigma = \sqrt{2\,609\,6399} = 51\,0846$$

Im Durchschnitt weichen die Umsätze um **51,08** EUR nach oben und unten vom Mittelwert ab.

Varianz und Standardabweichung sind die wichtigsten Streuungsmaße, die zwar umständlich zu berechnen sind aber den Vorteil haben, dass sie von allen Merkmalswerten abhängig sind und von Extremwerten nicht stark beeinflusst werden.

2.4 Der Variationskoeffizient

Sowohl die Standardabweichung als auch die mittlere Abweichung sind **absolute** Streuungsmaße, somit hängt ihr Wert von der Dimension der Merkmalswerte ab. Diese Streuungsmaße sind zum Vergleich verschiedener Grundgesamtheiten nicht geeignet.

Relative Streuungsmaße setzen die absolute Streuung ins Verhältnis zum Mittelwert. Das wichtigste relative Streuungsmaß, das von der Standardabweichung und dem arithmetischen Mittel ausgeht, ist der Variationskoeffizient.

Der Variationskoeffizient ist das Verhältnis der Standardabweichung zum arithmetischen Mittel. ausgedrückt in Prozentwerten.

$$v = \frac{\sigma}{x} \cdot 100 (\%)$$

Beispiel:

Der Vergleich des Durchschnittseinkommens der Bevölkerung zweier Länder führt zu folgendem Ergebnis:

$$\text{Land 1: } x_1 = 3\,200 \$ \quad \sigma_1 = 400 \$ \quad v_1 = 12,50 \%$$

$$\text{Land 2: } x_2 = 450 \$ \quad \sigma_2 = 100 \$ \quad v_2 = 22,22 \%$$

Obwohl im Land 1 die Standardabweichung größer ist als im Land 2, kann nicht gesagt werden, dass die Streuung der Einkommensverteilung des ersten Landes größer ist, da das Einkommensniveau beider Länder unterschiedlich ist.

Zum Vergleich eignet sich nur ein relatives Streuungsmaß. Die relative Streuung des Einkommens im Land 2 ist mit **22,22 %** größer als im Land 1.

Beispiel:

Die relative Streuung der Umsätze im Warenhaus "W" beträgt:

$$v = \frac{51,08}{95,79} \cdot 100 = 53,32 \%$$

C. Verhältnis- und Indexzahlen

1. Verhältniszahlen im Betrieb

Die Methoden der Darstellung des statistischen Materials dienen dazu, eine übersichtliche Form zu finden. Mittelwerte und Streuungsmaße haben die Aufgabe, die Struktur der Daten zu charakterisieren. Damit ist diesen beiden Methoden eine Reduktion bzw. Zusammenfassung der umfangreichen Daten auf einige wenige Zahlenwerte gemeinsam. Aufgabe der Verhältniszahlen ist es, die Beziehungen zwischen den einzelnen statistischen Merkmalen und Merkmalsausprägungen durchsichtig und übersichtlich zu machen, Vergleiche mit anderen Abteilungen oder Bereichen bzw. Vergleiche mit anderen Zeitpunkten oder Zeiträumen zu ermöglichen.

Bei der Bildung von Verhältniszahlen werden statistische Zahlen dadurch miteinander verglichen, dass man sie dividiert.

Es wird also ein Quotient (ein Verhältnis) zweier statistischer Zahlen gebildet. Man erhält die Aussage, in welchem Verhältnis die eine statistische Erscheinung zur anderen steht.

In der Praxis spricht man pauschal auch von Kennzahlen, wobei man diese Kennzahlen in

- Gliederungszahlen
- Beziehungszahlen
- Meßzahlen.

1.1 Gliederungszahlen

Bei den Gliederungszahlen wird eine Teilmasse einer ihr übergeordneten Gesamtmasse gegenübergestellt.

Gliederungszahl = Teilmasse * 100 / Gesamtmasse.

Die Größe im Zähler ist in der Größe des Nenners mit enthalten und stellt einen Teil desselben dar. Damit drückt die Gliederungszahl den Anteil der Teilmasse an der Gesamtmasse aus.

Beispiel:

Der Umsatz der Abteilung "Bekleidung" des Warenhauses "W" betrug im Gj. 1 insgesamt 8,6 Mio. EUR, der Gesamtumsatz wurde mit 29,2 Mio. EUR ausgewiesen.

Wie hoch ist der Anteil in Prozent am Gesamtumsatz?

? %	...	8,6 Mio. EUR
100 %	...	29,2 Mio. EUR

Anteil = $8,6 \cdot 100 / 29,2 = 29,45 \%$.

Die Gliederungszahl 29,45 gibt das Verhältnis des Teilumsatzes zum Gesamtumsatz an.

1.2 Beziehungszahlen

Bei den Beziehungszahlen werden verschiedenartige statistische Massen einander gegenübergestellt.

$$\text{Beziehungszahl} = \frac{\text{Gesamtmasse A}}{\text{Gesamtmasse B}} \cdot 100$$

Bekannte Beziehungszahlen sind die sogenannten Dichteziffern, z. B.:

$$\text{Eigenkapitalrentabilität ekr [\% p. a.]} = \frac{\text{Gewinn [EUR/Jahr]} \cdot 100}{\text{durchschnittl. Eigenkapital [EUR]}}$$

Auf jeden Fall ist vor der Berechnung von Beziehungszahlen darauf zu achten, dass die Definitionen und die Aussagekraft der Massen, die zueinander in Beziehung gesetzt werden, der Fragestellung entsprechen und zu sinnvollen Aussagen führen. Sinnvoll bedeutet in diesem Falle, dass eine sachliche Beziehung zwischen den Massen vorhanden sein muss.

1.3 Meßzahlen

Bei Meßzahlen werden gleichartige statistische Massen einander gegenübergestellt.

$$\text{Meßzahl} = \frac{\text{gleichartige Masse A}}{\text{gleichartige Masse B}} \cdot 100$$

Allerdings ist die Gleichartigkeit nicht eindeutig definiert. Man spricht jedoch dann von Gleichartigkeit, wenn es möglich ist, beide Massen zu einer übergeordneten Gesamtheit zusammenzufassen.

Beispiel:

Anzahl der leitenden Angestellten im Unternehmen betrage 25 Personen, die Gesamtzahl der Beschäftigten betrage 625 Personen.

Damit beträgt der Anteil der leitenden Angestellten an der Gesamtzahl der Beschäftigten 4 % ($25 \cdot 100 / 625 = 4 \%$).

Meistens dienen Meßzahlen zur Darstellung *zeitlicher Entwicklungen*; man spricht dann von sogenannten *dynamischen Meßzahlen*. Dabei werden gleiche Massen einander gegenübergestellt, wobei sie lediglich durch ihre zeitliche Entwicklung voneinander verschieden sind.

$$\text{Meßzahl} = \frac{\text{gleiche Masse zum Beobachtungszeitpunkt}}{\text{gleiche Masse zum Basiszeitpunkt}} \cdot 100$$

Die Berechnung einer dynamischen Meßzahl bezieht man immer auf ein Basisjahr, dessen Wert mit 100 % gleichgesetzt wird. Demgegenüber wird der prozentuale Wert des *Berichtsjahres* berechnet, wobei sich das Ergebnis als eine Steigerung oder eine Senkung in % ergibt. Die Wahl des Basis- und des Berichtsjahres hängt von der statistischen Fragestellung ab. Im volkswirtschaftlichen Vergleich wird die Forderung gestellt, dass das **Basisjahr** einen konjunkturrell normalen Verlauf haben muss.

Beispiel:

Der Umsatz eines Unternehmens im Geschäftsjahr Gj. 12 betrug 2.376 [TEUR], im Vorjahr Gj. 11 wurde ein Umsatz von 2.200 [TEUR] ausgewiesen. Der Umsatz stieg somit auf $2.376 \cdot 100 / 2.200 = 108 \%$.

Häufig werden *Meßzahlenreihen* erstellt, bei denen sich mehrere Meßzahlen auf dasselbe Basisjahr beziehen. Bei fortschreitenden Berechnungen wird das Basisjahr immer beibehalten. Treten jedoch im Laufe der Zeit *Strukturveränderungen* auf, so wird das Basisjahr als Vergleichsperiode ungeeignet. Für die Weiterführung der Reihe ist dann die Wahl eines *neuen Basisjahres* notwendig.

Die dynamischen Meßzahlen werden auch als einfacher, ungewogener Index bezeichnet.

2. Indexzahlen

Eine der wichtigsten Indexzahlen zur Charakterisierung der volkswirtschaftlichen Gesamtsituation ist der *Preisindex der Lebenshaltung*. Er ist eine Indexzahl und damit eng verwandt mit den dynamischen Meßzahlen.

Durch die Indexzahlen, auch den Preisindex der Lebenshaltung, wird eine zeitliche Entwicklung charakterisiert. Im Gegensatz zu den vorher besprochenen Meßzahlen wird aber nicht die zeitliche Entwicklung eines einzelnen Merkmalswertes, sondern die Entwicklung einer Vielzahl von Merkmalswerten aufgezeigt. Wie bei der dynamischen Meßzahl legt man auch hier ein Basis- bzw. Berichtsjahr fest und stellt einen Vergleich der Werte des Berichtsjahres mit denen des Basisjahres an. Von der Vielzahl der in die

Indexrechnung eingehenden Waren kann **jede einen anderen Preis** besitzen; es werden von den einzelnen Waren unterschiedliche Mengen verbraucht.

Damit ist der Index eine durchschnittliche zeitliche Preis- und Mengenentwicklung all dieser Waren.

2.1 Die Gewichtung

Bei der Berechnung der Indexzahl, die die durchschnittliche zeitliche Entwicklung einer Vielzahl von Tatbeständen charakterisiert, geht man davon aus, dass zunächst für jeden Tatbestand gesondert die dynamische Meßzahl berechnet wird. Anschließend wird aus allen Meßzahlen der Durchschnittswert gebildet.

Bei der Berechnung des einfachen Durchschnitts würden alle Waren gleichgewichtig in die Berechnung eingehen. Dies entspricht aber nicht der Wirklichkeit, da die Waren unterschiedliche wirtschaftliche Bedeutung haben (z. B. hat Lachs längst nicht die gleiche Bedeutung wie Brot) und deshalb entsprechend ihrem Rang im wirtschaftlichen System gewichtet werden müssen.

Man geht davon aus, dass Waren, die einen größeren Anteil am Gesamtumsatz haben, stärker berücksichtigt werden müssen, als Waren mit geringerem Anteil am Gesamtumsatz. dies bedeutet, dass die **Umsatzwerte** der einzelnen Waren als **Gewichtungsfaktoren** verwendet werden.

Folgende zwei Gewichtungssysteme haben sich durchgesetzt:

- Gewichte aus dem Basisjahr (*Index nach Laspeyres*)
- Gewichte aus dem Berichtsjahr (*Index nach Paasche*).

Für die Berechnung ergeben sich folgende Formeln, die man auch als aggregierte Indexformeln bezeichnet.

$$\text{Preisindex: } I_{o,i} = \frac{\sum p_i \cdot q_o}{\sum p_o \cdot q_o} \cdot 100$$

$$\text{Mengenindex: } I_{o,i} = \frac{\sum p_i \cdot q_o}{\sum p_o \cdot q_o} \cdot 100$$

Beispiel:

Ein Abteilungsleiter, der vier Warengattungen betreut, möchte wissen, wie sich die Preise und Mengen im Laufe von drei Perioden verändert haben.

Ware	Jahr 1		Jahr 2		Jahr 3	
	Preis in EUR p ₀	Menge in kg q ₀	Preis in EUR p ₁	Menge in kg q ₁	Preis in EUR p ₂	Menge in kg q ₂
A	2,20	560	2,00	580	2,80	600
B	14,00	112	12,00	148	15,00	150
C	40,00	18	40,00	20	50,00	22
D	9,60	680	9,40	720	9,80	730

Allgemeine Preisentwicklung in dem Zeitraum Jahr 1 (= Basis 0) bis Jahr 2 (= Berichtsperiode 1):

$$\begin{aligned}
 I_{0,1} &= \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} \\
 &= \frac{2,00 \cdot 560 + 12 \cdot 112 + 40 \cdot 18 + 9,40 \cdot 680}{2,20 \cdot 560 + 14 \cdot 112 + 40 \cdot 18 + 9,60 \cdot 680} \cdot 100 \\
 &= \frac{1120 + 1344 + 720 + 6392}{1232 + 1568 + 720 + 6528} \cdot 100 \\
 &= \frac{9576}{10048} \cdot 100 = 95,3
 \end{aligned}$$

Wie sich aus den Meßzahlen ergibt, wurde das Jahr 1 als Basis 0 mit 100 angenommen; im Vergleich zu dieser Basis ist das Preisniveau von der 0. zur 1. Periode auf 95,3 zurückgegangen.

Auf die gleiche Weise berechnet man den Preisindex für die Entwicklung im Zeitraum 0. bis 2. Periode:

$$\begin{aligned}
 I_{0,2} &= \frac{\sum p_2 \cdot q_0}{\sum p_0 \cdot q_0} \\
 &= \frac{10812}{10048} \cdot 100 = 107,60
 \end{aligned}$$

Gegenüber der Basisperiode ist das Preisniveau von 100 auf 107,6 gestiegen; damit ist der Preisrückgang in der Periode 1 ausgeglichen worden.

Da neben der Preisentwicklung auch die Mengenentwicklung berechnet werden soll, müssen die Preise der Basisperiode konstant gehalten werden. Die Menge hat sich von der Basisperiode 0 bis zu Periode 2 wie folgt entwickelt:

$$\begin{aligned} I_{0,2} &= \frac{\sum p_0 \cdot q_2}{\sum p_0 \cdot q_0} \cdot 100 \\ &= \frac{2,20 \cdot 600 + 14 \cdot 150 + 40 \cdot 22 + 9,60 \cdot 730}{2,20 \cdot 560 + 14 \cdot 112 + 40 \cdot 18 + 9,60 \cdot 680} \cdot 100 \\ &= \frac{1320 + 2100 + 880 + 7008}{1232 + 1568 + 720 + 6528} \cdot 100 \\ &= \frac{11308}{10048} \cdot 100 = 112,54 \end{aligned}$$

Die umgesetzte Menge stieg von der Basisperiode 0 bis zur Periode 2 von 100 auf 112,54. (Für Basisperiode 1 ergibt sich: 110,07).

Die Indexformeln nach Laspeyres können in der betrieblichen Praxis leicht angewandt werden, da sie den Vorteil haben, dass die Gewichte nur einmal bestimmt werden müssen, denn sie sind für jede Berechnung gleich. Selbst dann, wenn ein Zeitraum von mehreren Jahren verglichen wird, können die Gewichte beibehalten werden, wodurch sich der Rechenaufwand vermindert, denn der Nenner der Formel bleibt stets gleich.

Der wichtigste Vorteil der Laspeyreschen Formeln liegt in der direkten Vergleichbarkeit aller Zahlen einer Indexreihe, die nach dieser Formel bestimmt wurde.

Diese Vergleichbarkeit aller Werte resultiert aus dem festen Basisjahr, dessen Werte als Gewichte herangezogen werden, und damit aus der Konstanz der Gewichte.

Diesem Vorteil entspricht jedoch ein Nachteil. Bei der Verwendung konstanter Gewichte wird angenommen, dass sich nur ein Faktor ändert, aber der andere konstant bleibt. Dies bedeutet: Beim Preisindex wird angenommen, dass die Verkaufsmengen des Basisjahres sich in jedem Folgejahr nicht ändern; bei Mengenindex wird angenommen, dass sich die Preise im Laufe der Zeit nicht ändern. Diese Annahme ist jedoch höchst unrealistisch.

Besonders wenn die Berichtsperiode zeitlich weit von der Basisperiode entfernt ist, muss angenommen werden, dass der Laspeyres-Index unexakt ist, da die

unverändert beibehaltenen Gewichte der Basisperiode weder den Preiswandel noch die Mengenveränderungen erfassen.

2.3 Gewichte aus der Berichtsperiode - Index nach Paasche

$$\text{Preisindex: } I_{o,i} = \frac{\sum p_i \cdot q_i}{\sum p_o \cdot q_i} \cdot 100$$

$$\text{Mengenindex: } I_{o,i} = \frac{\sum p_i \cdot q_i}{\sum p_i \cdot q_o} \cdot 100$$

Beispiel:

Der Abteilungsleiter wählt eine andere Gewichtung für seine 4 Warengattungen; der Paasche-Index wird wie folgt gerechnet:

Preisindex:

$$\begin{aligned} I_{0,2} &= \frac{\sum p_2 \cdot q_2}{\sum p_o \cdot q_2} \cdot 100 \\ &= \frac{2,80 \cdot 600 + 15 \cdot 150 + 50 \cdot 22 + 9,80 \cdot 730}{2,20 \cdot 600 + 14 \cdot 150 + 40 \cdot 22 + 9,60 \cdot 730} \cdot 100 \\ &= \frac{12184}{11308} \cdot 100 = 107,75 \end{aligned}$$

Der Preisindex nach Laspeyres zeigt im vorhergehenden Beispiel einen Anstieg von der Periode 0 zur Periode 2 von 107,60, während der Preisindex nach Paasche einen Anstieg von 107,75 aufweist. Dieser Unterschied ist ein Indiz dafür, dass sich die Preis-Mengen-Struktur der Güter von der Basisperiode zur Berichtsperiode geändert hat. Wäre dies nicht der Fall, so würden beide Preisindizes den gleichen Wert aufweisen.

Mengenindex:

Bei der Berechnung der Mengenänderungen von der Periode 0 bis zur Periode 2 nach Paasche werden die Preise zum gegenwärtigen Zeitpunkt als Gewichte herangezogen:

$$\begin{aligned} I_{o,2} &= \frac{\sum p_2 \cdot q_2}{\sum p_2 \cdot q_o} \cdot 100 \\ &= \frac{12184}{10812} \cdot 100 = 112,69 \end{aligned}$$

Der Mengenindex nach Laspeyres beträgt in diesem Zeitraum 112,54 (vgl. das vorhergehende Beispiel), während der Mengenindex nach Paasche 112,69 beträgt. Auch diese Differenz weist darauf hin, dass eine Änderung in der Preis-Mengen-Struktur eingetreten ist.

Wie dieses Beispiel zeigt, hat der Preis- bzw. Mengenindex nach Paasche den Vorteil, daß er die Situation, die zum gegenwärtigen Zeitpunkt besteht, wiedergibt, er entspricht der Realität.

Allerdings stehen seiner Anwendung praktische Schwierigkeiten entgegen, denn beim Paasche-Index müssen sowohl der Preis als auch die Menge in der Berichtsperiode ermittelt werden.

Dies hat den Nachteil, dass der Arbeitsaufwand und damit auch die Kosten der Ermittlung höher liegen als bei Laspeyres-Index. Als weiterer Nachteil ergibt sich, dass Indizes nach Paasche keine durchlaufende Reihe sind und damit kein direkter Vergleich aller Indexwerte möglich ist.

Die ständige Neugewichtung führt zwar zu einer höheren Aktualität der repräsentativen Werte, doch sind zum Vergleich der Entwicklung über mehrere Perioden zusätzliche Berechnungen notwendig.

Ende des Skripts