

# 2

## Ein ganz durchschnittliches Kapitel

### *Die mathematische Definition des „Typischen“*

Gewöhnlich gebrauchen wir die Begriffe „Durchschnitt“ oder „Mittelwert“ ganz locker, und wir wissen bis zu einem bestimmten Grad, was wir damit meinen. Wir beziehen uns auf etwas, was der „Norm“ entspricht bzw. „typisch“ ist und für die Mehrheit der Leute gilt. Oder wir denken an den gesunden Menschenverstand und an mittlere, ausgewogene Umstände und Erwartungen. Wir können Mittelwerte auch in einer etwas geringschätzigen Weise auffassen wie in den Formulierungen „nur Durchschnitt“ oder „nur Mittelmaß“. Sagt man von einem Sportler, er zeige durchschnittliche Leistungen, so ist das nicht gerade eine Schande, aber auch kein Grund, die Korken knallen zu lassen.

Aber inwieweit bringt uns dieser Gebrauch der Begriffe in der Alltagssprache weiter? Was passiert, wenn wir eine Zahl für den Mittelwert angeben müssen? In vielen Bereichen des modernen Lebens müssen wir aus einem Berg von Daten einen „typischen“ Wert heraussuchen. An diesem Wert können wichtige Entscheidungen hängen (siehe Kap. 15). Das ist insbesondere in der Statistik und Ökonomie wichtig, wo die Analysten nach einfachen Wegen

suchen, um komplizierte Zusammenhänge zu beschreiben. Angefangen bei den immensen staatlichen Sozialausgaben bis zur Kündigung und Einstellung von Mitarbeitern werden Prioritäten oft aufgrund von Forschungen und Trends gesetzt, die von der Definition eines „Mittelwertes“ abhängig sind. Überall um uns herum hören wir im politischen und wirtschaftlichen Diskurs Argumente, in denen von „Durchschnittseinkommen“, „durchschnittlicher Lebenserwartung“, „durchschnittlicher Rückfallrate“ oder vom „Notendurchschnitt“ in der Schule die Rede ist.

Natürlich ist hier die Mathematik die Grundlage für alles. Die mathematischen Methoden, um einen „typischen“ Wert zu definieren, gehen bis auf Pythagoras und seine Nachfolger im alten Griechenland zurück. Seit damals wurde eine ganze Reihe von Methoden entwickelt, zu einem „Mittelwert“ zu kommen. Wie es aussieht, ist der Oberbegriff „Mittelwert“ (oder auch „Durchschnitt“) zu verschwommen, und man sollte besser von Bereichsmittel, Median, Modus und von arithmetischem oder geometrischem Mittel sprechen.

## **Zwischen den Extremen: das Bereichsmittel**

Die schnellste und einfachste Methode, um zu einem Mittelwert zu kommen, ist die Berechnung der Mitte zwischen den Extremen. Der kleinste erwachsene Mensch, den es je gab, ist Chandra Bahadur Dangi mit ganzen 54,6 cm, der größte war Robert Wadlow mit 272 cm. Um einen Mittelwert der Größe des Menschen unserer Tage zu berechnen, addiert man also die beiden Extremwerte, dividiert durch 2 und erhält 163,3 cm.

Offensichtlich ist dieser Mittelwert, das sogenannte Bereichsmittel, ein allzu grobes Maß. Seine Vor- und Nachteile können wir auch an unserem Beispiel erkennen: Der Vorteil ist, dass der Wert schnell und einfach zu bestimmen ist. Der Nachteil ist, dass er auf den Extremwerten beruht und damit auf den *untypischsten* Werten.

1890 fand man in einem Grab aus der Bronzezeit im französischen Castelnau-le-Lez einen riesigen menschlichen Armknochen und einige Beinknochen. Sie schienen von einem Menschen zu stammen, der ungefähr 3,5 m groß war. Hätte der Fund bestätigt werden können, wäre der „Riese von Castelnau“ sicher eine bemerkenswerte Entdeckung gewesen. Der Fund hätte aber auch die Schwächen des Bereichsmittels unterstrichen, denn die Durchschnittsgröße des Menschen wäre plötzlich auf 202,3 cm angewachsen.

Trotz dieser Mängel war diese Mittelwertbestimmung lange Zeit die gängigste Methode. So wollte beispielsweise der Astronom und Mathematiker Ptolemäus den durchschnittlichen Stand der Sonne über dem Horizont zur Mittagszeit wissen. Um diesen Wert zu erhalten, nahm er den Minimalwert am längsten Tag des Jahres, addierte ihn zu dem Maximalwert am kürzesten Tag und erhielt als Mittelwert des Abstands vom Zenit etwa  $23,5^\circ$  und damit den Wert, der die Neigung der Erdachse angibt.

## Der große Auftritt in der Mitte: der Median

Eine andere Art von Mittel, dessen Grundprinzip ähnlich leicht zu begreifen ist, stellt der Median dar. Seit dem späten 19. Jahrhundert wurde er in zunehmenden Maß in der Na-

turwissenschaft verwendet. Man reiht alle fraglichen Daten der Größe nach auf und erhält als Median den Wert, der in der Mitte der Reihe steht. Hat man beispielsweise fünf Hunde, die 1, 2, 4, 8 und 13 Jahre alt sind, ist der Median 4. 50 % der Hunde sind jünger, 50 % älter.

Obwohl das ein vernünftiger Ansatz zu sein scheint, wird der Median erst seit relativ kurzer Zeit ernsthaft in wissenschaftlichen Kreisen verwendet. Der Psychologe Gustav Fechner (1801–1887) war einer der Ersten, der mit ihm arbeitete, als er in der Mitte des 19. Jahrhunderts eine Analyse der Farbinterpretation des Menschen anstellte. Fechner fand heraus, dass eine bestimmte Eigenschaft des Medians sehr nützlich ist: Er vermindert den Abstand zu den einzelnen Datenpunkten. Um das zu demonstrieren, gehen wir zu den fünf Hunden zurück, deren Median 4 ist. Berechnen wir nun die Differenz zwischen dem Alter von jedem Hund und dem Median, erhalten wir 3, 2, 0, 4 und 9. Die Summe dieser Differenzen ist 18.

Führen wir diese Rechnung mit irgendeinem anderen Wert anstelle des Medians durch, etwa mit 5, erhalten wir die Differenzen 4, 3, 1, 3 und 8 und die Summe 19. Es ist kein Zufall, dass die Summe größer ist als die in Bezug zum Median: Mathematische Überlegungen zeigen, dass das *immer* so ist.

## Der häufigste Wert: der Modus

Der Modus ist eine weitere klare und einfache Definition von „Mittelwert“. In diesem Fall ist Mittelwert mit „häufigster Wert“ gleichbedeutend. In einer Schulklasse sind sie-

ben Kinder 8 Jahre alt, fünfzehn sind 9 und zwei sind 10. In diesem Fall ist der Modus 9. Der Modus funktioniert nur gut, wenn die Zahl der möglichen Werte begrenzt ist. So kann der Modus beispielsweise dazu dienen, das Ergebnis von Wahlen zu bestimmen, bei denen die Wähler unter einer kleinen Zahl von Kandidaten wählen können (siehe Kap. 10). Anders als die übrigen Mittelwerte kann man den Modus auch bei Problemen bestimmen, bei denen es nicht um Zahlen geht.

Der Statistiker Wilson Allen Wallis (1912–1998) hat von einem Beispiel für den Gebrauch des Modus bei dem Historiker Thukydides im 5. Jahrhundert vor unserer Zeitrechnung berichtet. Er beschreibt den Peloponnesischen Krieg zwischen Athen und Sparta. Die Verbündeten der Athener wurden in Plataiai von den Spartanern belagert und wollten die Höhe der Befestigungsanlagen wissen, die diese um ihre Stadt herum aufgebaut hatten, denn sie planten, sie zu übersteigen und zu entkommen. Um dafür Leitern der richtigen Länge bauen zu können, zählten sie die Zahl der Ziegelschichten der Befestigungsbauten. Ihnen war klar, dass jeder, der die Schichten zählte, einen Fehler machen konnte, deshalb ordnete der Befehlshaber an, dass viele Menschen gleichzeitig zählten. Sie verglichen dann ihre Ergebnisse und wählten die häufigste Zahl, also den Modus, als die richtige Antwort aus.

Die Idee des Kommandeurs von Plataiai ist ein Beispiel für etwas, was wir heute „Schwarmintelligenz“ nennen würden. Solche Verfahren bewähren sich auch in weit weniger lebensbedrohenden Situationen. Ein Beispiel ist, die Anzahl der Bonbons in einem großen Glas zu schätzen. Vermutlich ist jede Einzelschätzung falsch, aber Tests haben gezeigt,

dass die Abfrage einer großen Anzahl individueller Schätzungen zu einem vernünftigen, mittleren Wert führt, weil sich Über- und Unterschätzungen gegenseitig aufheben. Es kann sich dabei um den Median handeln, aber auch der Modus und das arithmetische Mittel erfüllen ihren Zweck.

## **Der moderne Mittelwert: das arithmetische Mittel**

Der bekannteste Mittelwert der modernen Zeit ist das arithmetische Mittel. (Zur „Regression zur Mitte“ siehe Kap. 18.) Auch dieses Mittel ist klar definiert und umfasst Additionen und eine Division. Wollen wir das mittlere Gewicht (genauer gesagt: die mittlere Masse) von sechs verschieden großen Eiern wissen, addieren wir die jeweiligen Gewichte – 50, 52, 58, 61, 63 und 70 g – und dividieren die Summe (354 g) durch die Zahl der Eier, also sechs. Wir erhalten den Mittelwert 59 g.

Redet heute jemand vom Mittelwert, meint er höchstwahrscheinlich das arithmetische Mittel. Diese Vorherrschaft reicht bis in das frühe 17. Jahrhundert zurück. Das arithmetische Mittel wurde wie zuvor der Modus verwendet, um die Abweichungen der einzelnen Werte vom Mittel möglichst klein zu halten. Diesmal ging es um die Seefahrt. Um die Fahrtrichtung zu bestimmen, benützte der Steuermann einen Kompass, der die Nordrichtung angab. Die frühen Modelle dieses Instruments wiesen aber große Fehler auf, insbesondere bei stürmischer See. So entwickelte sich als Standardverfahren, mehrere Ablesungen zu machen und dann den Mittelwert als den „wahren“ Wert zu neh-

men – vielleicht, nachdem man zuvor alle Werte verworfen hatte, die allzu weit abwichen.

Für unsere Augen erscheint es seltsam, dass das arithmetische Mittel so lange brauchte, um anerkannt zu werden, nachdem es schon 2000 Jahre zuvor von den pythagoreischen Mathematikern sorgfältig studiert worden war. Die Pythagoreer sahen in ihm aber weniger ein statistisches Werkzeug, sondern schätzten seine „reinen“ mathematischen Eigenschaften und seine Anwendungsmöglichkeit in der Musiktheorie. Das berühmteste Ergebnis der Untersuchungen der Pythagoreer war die Beziehung zwischen dem arithmetischen Mittel und einer anderen wichtigen Art von Mittelwert: dem geometrischen Mittel.

## Durchschnittlich interessant: das geometrische Mittel

Ein Ausflug ins Reich der Zinssätze hilft uns, die Beziehung zwischen arithmetischem und geometrischem Mittel zu verstehen. Angenommen, wir investieren Geld, und die Anlage entwickelt sich gut. Im ersten Jahr steigt der Wert um 50 %, im zweiten um 10 % und im dritten um 15 %. Wie berechnen wir die durchschnittliche Wertsteigerung in den drei Jahren? Hier sind weder Bereichsmittel noch Modus noch Median geeignet, deshalb versuchen wir es mit dem arithmetischen Mittel. Wir addieren die Prozentzahlen, teilen die Summe durch die Zahl der Jahre (3) und erhalten 25 %.

Damit *scheint* alles klar, aber wenn wir genauer hinschauen, ist die Antwort doch nicht in Ordnung. Angenommen,

unser Startkapital war 100 €. Nach dem ersten Jahr ist es auf 150 € angewachsen. Nach dem zweiten Jahr hat es um 10 % der 150 € zugenommen, also um 15 € auf 165 €, nach dem dritten Jahr dann um 15 % der 165 €, also um  $0,15 \cdot 165 \text{ €} = 24,75 \text{ €}$ , sodass der Endbetrag nun 189,75 € beträgt.

Wenn nun unsere Rechnung mit dem arithmetischen Mittel richtig ist, müssten wir zum gleichen Endbetrag kommen. Aber 25 % auf die anfänglichen 100 € bringen im ersten Jahr 125 €, im zweiten dann mit 25 % der 125 € den Betrag 156,25 € und im dritten Jahr weitere 25 % dieses Betrags, also schließlich 195,31 €. Das ist also deutlich mehr.

Unser Beispiel einer Geldanlage unterscheidet sich in einem Kriterium deutlich von unseren Eiern. Bei den Eiern machte es Sinn, einfach die Einzelgewichte zu addieren. Bei den Prozentzahlen der Geldanlage macht das dagegen wenig Sinn, da sich Prozentzahlen auf Multiplikationen stützen und nicht auf Additionen. Einen Geldbetrag um 15 % zu steigern bedeutet, ihn mit 1,15 zu multiplizieren. Um also die Steigerung nach drei Jahren zu berechnen, müssen wir die 100 € nacheinander mit den drei Prozentzahlen multiplizieren:  $100 \text{ €} \cdot 1,5 \cdot 1,1 \cdot 1,15 = 189,75 \text{ €}$ . Wir müssen also den ursprünglichen Betrag mit 1,8975 multiplizieren, um zum richtigen Resultat zu kommen.

In unserem Beispiel mit den drei Jahren erhält man den Durchschnittszins  $a$ , indem man die Kubikwurzel aus 1,8975 zieht, mathematisch formuliert:  $\sqrt[3]{1,8975}$ . Das Ergebnis ist 1,23802. Jetzt haben wir die richtige Lösung. Ein Anfangskapital von 100 €, das mit 23,802 % verzinst wird, liefert nach drei Jahren 189,75 €.



Wenn wir unser Beispiel verallgemeinern, ist das *arithmetische* Mittel von  $n$  Zahlen deren Summe, dividiert durch die Anzahl  $n$ . Das *geometrische* Mittel ist hingegen die  $n$ -te Wurzel aus dem Produkt der  $n$  Zahlen. Wenn  $y$  das geometrische Mittel und  $x$  das Produkt der  $n$  Zahlen ist, gilt  $y = \sqrt[n]{x}$  und umgekehrt  $x = y^n$ , was voll ausgeschrieben so aussieht:

$$x = \underbrace{y \cdot y \cdot \dots \cdot y}_{n \text{ Mal}}.$$

So ist beispielsweise  $\sqrt[4]{16} = 2$ , weil  $2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$  ist, während  $\sqrt[3]{125} = 5$  ist.

## Andere Fragen, andere Antworten

Für Ökonomen, deren Aufgabe es ist, Inflationsraten und mittlere Preissteigerungen abzuschätzen, kann das Theorem des Pythagoras über den Unterschied der beiden Mittelwerte (siehe Kasten) ganz wesentliche Auswirkungen haben. An erster Stelle stehen bei den Ökonomen oft Preisindizes oder Indizes der Lebenshaltungskosten, bei denen es darum geht, in bestimmten zeitlichen Abständen einen typischen „Warenkorb“ zu beobachten und zu analysieren, ob die Preise steigen oder fallen. In Großbritannien gibt es zwei derartige Indizes, den Preisindex (RPI: Retail Price Index) und den Lebenshaltungsindex (CPI: Consumer Price Index). Der RPI enthält die Wohnkosten, der CPI nicht. Der wesentliche *mathematische* Unterschied ist aber, dass der RPI auf dem arithmetischen Mittel basiert, der CPI auf dem geometrischen. Jeder Vergleich der beiden Indizes führt zu dem nüchternen mathematischen Resultat, dass

das arithmetische Mittel von Daten immer größer ist als das geometrische Mittel. Jede Art von Ausgabe oder jede Investition, die mit dem einen oder anderen Index verbunden ist, spiegelt diesen Unterschied wider.

Die meisten Briten hätten es lieber, wenn ihre Rente oder Pension an den RPI geknüpft wäre. Inhaber einer Firma würden jedoch ihre Entscheidungen über Lohnerhöhungen lieber an den CPI knüpfen. Wenn Regierungen und Organisationen vom einen zum anderen Index wechseln, ist das stets ein Grund für Misstrauen. Man sollte dann die Motive hinterfragen, denn meistens wird zu dem anderen Index gewechselt, um bestimmte statistische Aussagen zu erhalten. (Der deutsche Verbraucherpreisindex VPI enthält auch die Wohnkosten und basiert auf dem arithmetischen Mittel.)

Das alles soll nur zeigen, dass es bei den Mittelwerten (wie auch in der Wissenschaft, ja wie im Leben überhaupt) nicht nur um die richtigen Antworten geht, sondern vor allem um die richtigen Fragen. Die Wahrheit ist, dass es für einen Datensatz nicht *den* einen und einzigen Mittelwert gibt. Bevor wir also den geeignetsten Mittelwert wählen, müssen wir überlegen, was wir mit dieser Zahl anfangen wollen, was für das aktuelle Szenario wichtig ist (und was nicht) und wie andere den Mittelwert interpretieren werden. Ganz allgemein gesagt kann uns die Mathematik sehr viel über die Welt sagen, aber man darf ihre Formeln nicht blind anwenden, ohne dabei den Kontext zu berücksichtigen.

<http://www.springer.com/978-3-642-41792-4>

Das Chaos im Karpfenteich oder Wie Mathematik  
unsere Welt regiert

Elwes, R.

2014, XI, 464 S. 51 Abb., Softcover

ISBN: 978-3-642-41792-4