

Kapitel 2.

Mittelwerte

Im Zusammenhang mit dem Begriff der Verteilung, der im ersten Kapitel eingeführt wurde, taucht häufig die Frage auf, wie man die vorliegenden Daten durch eine geeignete Größe „repräsentieren“ kann. Dabei werden oft Begriffe wie Mittelwert oder Durchschnitt verwendet, die einem bereits aus der Umgangssprache bekannt sind. Bei unserem Kaffeehaus-Beispiel wäre etwa die Frage naheliegend, wie viel eine Melange im Durchschnitt kostet. Gesucht ist hier also nichts anderes als eine quantitative Zusammenfassung der Daten, was zu den typischen Aufgaben der Statistik gehört. Mit diesem aus praktischer Sicht sehr wichtigen Thema werden wir uns auf den folgenden Seiten näher beschäftigen. Wir werden dabei verschiedene Maßzahlen kennenlernen, die auf unterschiedliche Weise der Vorstellung von einem Mittelwert oder Durchschnitt entsprechen. Etwas allgemeiner spricht man in diesem Zusammenhang auch von Lagemaßzahlen. Diese stellen eine Form von komprimierter Information dar. Außerdem erleichtern sie die Vergleichbarkeit von Daten für den Fall, dass nicht nur ein, sondern mehrere Datensätze vorliegen. Wenn man zum Beispiel an den Preis von Produkten wie Kaffee, Bier, Brot, Obst oder auch Benzin denkt, dann hat man es ja nicht mit Einheitspreisen, sondern mit einer Vielzahl von Einzelpreisen zu tun. Wenn daher in Medienberichten vom aktuellen Benzinpreis oder dem diesjährigen Bierpreis am Münchner Oktoberfest die Rede ist, dann sind natürlich Durchschnittspreise gemeint, auch wenn dies nicht immer eigens erwähnt wird.

Betrachten wir ein anderes Beispiel. Abbildung 2.1 zeigt die jährlichen Wachstumsraten des Bruttoinlandsprodukts (BIP) für die EU, Deutschland, Österreich und die Schweiz. Eine naheliegende Frage in diesem Fall wäre wohl die nach einer geeigneten Maßzahl, um das durchschnittliche jährliche Wachstum zu beschreiben. Dies würde nicht nur die Vergleichbarkeit der einzelnen Zeitreihen erleichtern. Es könnte nämlich auch von Interesse sein, das jeweilige Wachstum für verschiedene Abschnitte einer gegebenen Zeitreihe miteinander zu vergleichen. Wir

Kapitel 2. Mittelwerte

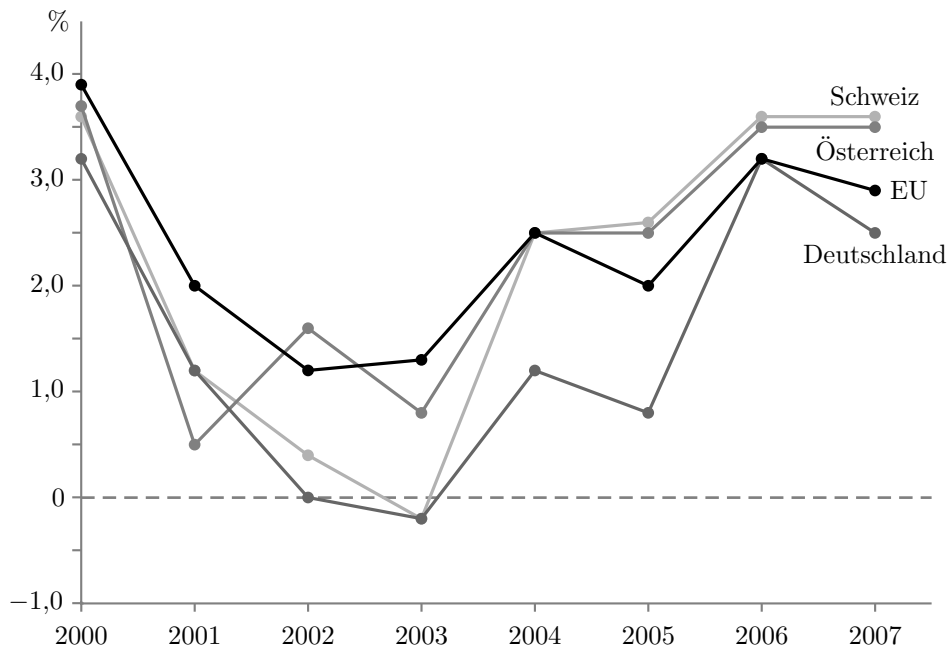


ABBILDUNG 2.1. BIP-Wachstumsraten für die EU, Deutschland, Österreich und die Schweiz, 2000–2007 (Jahresdaten, real, in %)

werden sehen, dass die Beschreibung des Wachstums einer Zeitreihe ein spezielles Mittelwertkonzept erfordert, das man auch in der Finanzmathematik verwendet.

Falls nicht anders vermerkt, setzen wir ab jetzt voraus, dass die verwendeten Ausprägungen immer von quantitativen Merkmalen stammen und dass man uneingeschränkt mit ihnen operieren kann. Beobachtungen, das heißt beobachtete Daten, werden dabei üblicherweise symbolisch mit x_1, x_2, \dots, x_n bezeichnet.

Es folgt ein kurzer Überblick über den Inhalt dieses Kapitels. Der erste Abschnitt beschäftigt sich mit dem am häufigsten verwendeten Mittelwert, dem arithmetischen Mittel. Dabei werden wir neben der traditionellen Variante auch eine Variante präsentieren, bei der die einzelnen Werte eine Gewichtung aufweisen können. Im zweiten Abschnitt wird das geometrische Mittel behandelt, das typischerweise bei Zeitreihendaten verwendet wird, wenn es darum geht, so etwas wie eine durchschnittliche Wachstumsrate zu bestimmen. Im nächsten Abschnitt wird kurz auf einige weitere Maßzahlen eingegangen, die ebenfalls im Zusammenhang mit Mittelwerten relevant sind, nämlich der Modus, der Median bzw. allgemeiner die Quantile. Eine spezielle grafische Darstellungsform, die Boxplots, wird im vierten Abschnitt vorgestellt. Im letzten Abschnitt schließlich werden die vier Aktienindizes an Hand der durchschnittlichen Wachstumsrate bzw. Rendite miteinander verglichen.

2.1. Arithmetisches Mittel

Wann immer die Frage nach einem Durchschnitt, einem durchschnittlichen Wert oder Mittelwert gestellt wird, wird man häufig zunächst an das arithmetische Mittel denken¹. Es ist die mit Abstand populärste Lagemaßzahl und wird üblicherweise durch einen Querstrich, zum Beispiel \bar{x} , gekennzeichnet².

Arithmetisches Mittel

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

Das arithmetische Mittel erhält man also dadurch, indem die Summe der Daten durch ihre Anzahl dividiert wird. Bei den gesamten Melange-Preisen ergibt sich daher für das arithmetische Mittel der Wert

$$\bar{x} = \frac{3,10 + 2,40 + \cdots + 4,40 + 2,60}{70} = 3,10$$

Die Frage nach dem durchschnittlichen Preis für eine Melange wäre in diesem Fall mit der Angabe des Wertes $\bar{x} = 3,10$ beantwortet. Berechnet man noch separat die arithmetischen Mittel der Datensätze für den 1. Bezirk bzw. für die übrigen Bezirke, dann erhält man die Werte $\bar{x}_1 = 3,40$ bzw. $\bar{x}_2 = 2,85$, was erwartungsgemäß den Unterschied im Preisniveau zwischen dem Stadtzentrum und den übrigen Bezirken widerspiegelt.

Das arithmetische Mittel kann man vergleichen mit dem Gleichgewichtspunkt einer Waage. Betrachtet man die Abweichungen der Beobachtungen vom arithmetischen Mittel, das heißt die Differenzen $x_i - \bar{x}$, dann zeigt sich, dass sich negative und positive Abweichungen gegenseitig aufheben. Es gilt nämlich:

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

Durch diese Eigenschaft wird das arithmetische Mittel charakterisiert. Mit anderen Worten, bei gegebenen Beobachtungen x_1, x_2, \dots, x_n ist das arithmetische Mittel \bar{x} gerade diejenige Zahl, für die die Summe der Abweichungen den Wert Null ergibt.

Fasst man die 70 Einzelwerte in einer Häufigkeitstabelle zusammen, wie wir es bereits in Kapitel 1 getan haben, dann wird die Darstellung der Melange-Preise

TABELLE 2.1. Häufigkeiten der
Melange-Preise

Preis	absolute Häufigkeit	Preis	absolute Häufigkeit
2,30	2	3,40	3
2,40	4	3,50	3
2,50	2	3,60	3
2,60	9	3,70	3
2,70	3	3,80	0
2,80	2	3,90	0
2,90	8	4,00	1
3,00	8	4,10	0
3,10	6	4,20	1
3,20	4	4,30	0
3,30	4	4,40	4

nicht nur übersichtlicher, sondern erlaubt auch eine alternative und elegantere Berechnung des arithmetischen Mittels. Das arithmetische Mittel lässt sich nämlich jetzt auch so berechnen:

$$\bar{x} = \frac{2,30 \cdot 2 + 2,40 \cdot 4 + \dots + 4,40 \cdot 4}{70} = 3,10$$

Dabei wird jeder Preis mit der Häufigkeit seines Auftretens multipliziert („gewichtet“). Bezeichnet man allgemein die verschiedenen Merkmalsausprägungen mit a_1, a_2, \dots, a_k , dann erhält man also das arithmetische Mittel, indem die einzelnen Merkmalsausprägungen mit den entsprechenden absoluten Häufigkeiten h_1, h_2, \dots, h_k gewichtet werden. Auf diese Weise erhält man formal das sogenannte gewichtete arithmetische Mittel.

Gewichtetes arithmetisches Mittel

$$\bar{x} = \frac{a_1 \cdot h_1 + \dots + a_k \cdot h_k}{n} = \frac{1}{n} \sum_{i=1}^k a_i \cdot h_i$$

Diese Darstellung des arithmetischen Mittels ermöglicht übrigens eine Modifikation, bei der man statt der absoluten Häufigkeiten spezielle positive Gewichte g_1, g_2, \dots, g_k verwenden kann, wobei dann der Zähler durch die Summe der

Gewichte dividiert wird. Auf diese Weise könnte man die einzelnen Merkmalsausprägungen quasi mit ihrer „Bedeutung“ gewichten. Man denke etwa an die Gewichtung von Punktezahlen bei Prüfungen mit unterschiedlichem Schwierigkeitsgrad.

Wenn man das mit den absoluten Häufigkeiten berechnete arithmetische Mittel etwas umformt, dann zeigt sich, dass man \bar{x} darstellen kann als die Summe der Produkte aus den verschiedenen Merkmalsausprägungen und den zugehörigen relativen Häufigkeiten

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k a_i \cdot h_i = \sum_{i=1}^k a_i \cdot \frac{h_i}{n} = \sum_{i=1}^k a_i \cdot f_i$$

Auf diese Darstellung werden wir in der Wahrscheinlichkeitsrechnung noch einmal zurückkommen. Mit ihrer Hilfe lässt sich nämlich der Begriff des Erwartungswerts einer Zufallsvariablen motivieren.

klausurtipp

Falls man das gewichtete arithmetische Mittel verwendet, sollte man sich klar darüber sein, über welche Werte eigentlich „gemittelt“ wird. Hier kommt es immer wieder zu Verwechslungen. Betrachten wir zum Beispiel die folgende Berechnung des arithmetischen Mittels der Ausprägungen 1, 2, 3 und 4 eines Merkmals mit den absoluten Häufigkeiten 5, 3, 4 und 8:

$$\bar{x} = \frac{1 \cdot 5 + 2 \cdot 3 + 3 \cdot 4 + 4 \cdot 8}{10} = 3,5$$

Dieses Ergebnis ist natürlich falsch! In diesem Fall ist nämlich der Nenner nicht gleich der Summe der Häufigkeiten, sondern gleich der Summe der vier verschiedenen Ausprägungen, die tatsächlich den Wert 10 hat. Wenn man will, könnte man sagen, dass hier das gewichtete arithmetische Mittel von vier Zahlen berechnet wurde, wobei die „Ausprägungen“ 5, 3, 4 und 8 mit den „Häufigkeiten“ 1, 2, 3 und 4 gewichtet wurden. Das war aber nicht gefragt!

Obwohl das arithmetische Mittel derjenige Mittelwert ist, der am häufigsten verwendet wird, ist seine Anwendung nicht immer ganz unproblematisch. Da nämlich bei seiner Berechnung sämtliche Beobachtungen berücksichtigt werden, können zum Beispiel extreme Werte, sogenannte Ausreißer, den Mittelwert unter Umständen stark beeinflussen. Das ist nicht unbedingt eine wünschenswerte Eigenschaft, wenn man an einem repräsentativen Mittelwert interessiert ist. Auf diesen Aspekt werden wir bei der Definition des Medians noch eingehen.