Alt

Statistik

Eine Einführung für Wirtschaftswissenschaftler

Linde international

2., aktualisierte Auflage

Kapitel 1.

Einführung

„*Ein Gast sitzt im Kaffeehaus und trinkt Kaffee.*“ Dieser Satz aus einer berühmten Abhandlung[1] des Schriftstellers, Kritikers und Publizisten *F. Torberg* (1908–1979) wird von seinem Autor zum Anlass genommen, eine Reihe von Fragen zu erörtern, die die nach seiner Ansicht komplizierteste der funktionierenden Legenden Wiens betreffen – das Wiener Kaffeehaus. Die drei wichtigsten Fragen für ihn sind dabei: Erstens „*Wer ist der Gast?*“, zweitens „*In welcher Art von Kaffeehaus sitzt er?*“ und drittens „*Was ist es für ein Kaffee, den er trinkt?*“. Für Studierende der Wirtschaftswissenschaften dürfte es nahe liegend sein, diesen Fragen unmittelbar eine weitere folgen zu lassen, nämlich: „*Was kostet der Kaffee?*“. Und das wird auch die erste in einer Reihe interessanter Fragen sein, mit denen wir uns in dieser Einführung in die Statistik beschäftigen werden.

Damit sind wir beim Thema dieses Buches angelangt. Einfach ausgedrückt könnte man sagen, dass sich Statistik vor allem mit der Zusammenfassung von Daten beschäftigt. Diese Beschreibung macht deutlich, was im Mittelpunkt der Statistik steht, nämlich Daten. Um es gleich vorwegzunehmen, Daten und Zahlen sind nicht dasselbe! Daten werden zwar häufig zahlenmäßig erfasst, allerdings stehen sie immer in einem sachlichen Kontext. Bei den in einer Klausur erzielten Punktezahlen kann man daher von Daten sprechen, bei beliebig vorgegebenen Zahlen wie 2, 5, 7, 10 dagegen nicht.

Angesichts einer immer komplexer werdenden Welt mit ihrem wachsenden Bedürfnis nach Daten ist es kaum überraschend, dass statistische Themen in Beruf und Alltag mehr und mehr an Bedeutung gewinnen, ob es nun um die aktuelle Preisentwicklung, die Höhe des Wirtschaftswachstums oder der Arbeitslosenrate geht oder um den Ertrag und das Risiko einer Finanzanlage. Im Laufe dieser Einführung wird man dabei auch die Erfahrung machen, dass es oft verschiedene Möglichkeiten zur Beantwortung einer statistischen Frage bzw. zur Lösung eines statistischen Problems gibt, die jede für sich einen anderen Aspekt betont. Dieser Umstand hat der Statistik gelegentlich den Vorwurf eingebracht, dass man mit ihrer Hilfe letztlich alles beweisen kann.

TABELLE 1.1. Preise für eine Melange in Wiener Kaffeehäusern[2] (in Euro)

1. Bezirk									
3,10	2,40	3,00	3,20	3,70	4,00	3,20	3,40	2,80	3,20
3,50	3,40	3,10	4,20	3,30	3,00	4,40	3,10	2,90	4,40
3,60	3,30	3,10	2,90	4,40	3,60	3,60	3,50	3,50	3,40
Übrige Bezirke									
2,30	3,00	2,90	2,60	3,00	2,90	2,60	2,90	2,90	2,70
2,60	2,90	2,50	3,20	2,80	3,70	2,40	3,00	2,70	3,00
3,00	3,00	2,60	2,40	3,30	2,90	3,10	2,60	3,70	2,60
3,30	3,10	2,50	2,60	2,70	2,40	2,60	2,30	4,40	2,60

Um die Darstellung etwas konkreter zu machen, werden wir am Beginn einen einfachen Datensatz verwenden, anhand dessen wir verschiedene statistische Fragestellungen behandeln werden. Bei diesem Datensatz, der in Tabelle 1.1 wiedergegeben ist, handelt es sich um Preise für eine Tasse Kaffee (genauer: für eine Wiener Melange), die für 70 Wiener Kaffeehäuser erhoben wurden[3]. Wir werden dabei versuchen, diesen Datensatz „statistisch“ etwas aufzubereiten. Dazu gehört insbesondere die Darstellung der Daten mit Hilfe von Tabellen bzw. Grafiken, aber natürlich auch der Hinweis auf verschiedene Möglichkeiten, wenn es zum Beispiel um die Berechnung eines geeigneten Mittelwerts oder einer Maßzahl zur Beschreibung der Streuung der Preise geht.

Es folgt ein kurzer Überblick über den Inhalt dieses Kapitels. Der erste Abschnitt beschäftigt sich mit einem elementaren statistischen Konzept – der Grundgesamtheit (die manchmal auch als Population bezeichnet wird). Weitere wichtige Konzepte werden im zweiten Abschnitt behandelt, nämlich Merkmale (Variablen) und ihre Verteilungen. Dabei wird insbesondere auf die Unterschiede zwischen quantitativen und qualitativen Merkmalen bzw. zwischen diskreten und stetigen Merkmalen eingegangen. Im dritten Abschnitt werden anhand verschiedener Beispiele zwei Darstellungsformen betrachtet, die typisch für den Umgang mit statistischen Daten sind, nämlich Tabellen und Grafiken. In diesem Zusammenhang wird auch eine Kombination dieser beiden Darstellungsformen erwähnt. Der vierte Abschnitt präsentiert den ersten Teil einer kleinen empirischen Fallstudie. Dabei werden Eigenschaften verschiedener Aktienindizes aus Deutschland (DAX), Österreich (ATX), der Schweiz (SMI) und den USA (Dow-Jones-Index) untersucht. Diese Fallstudie wird in den folgenden fünf Kapiteln fortgesetzt, wobei immer ein enger thematischer Zusammenhang zum jeweiligen Kapitel besteht.

1.1. Die Grundgesamtheit

Wenn man im Rahmen der Mengenlehre in einem bestimmten Zusammenhang verschiedene Mengen betrachtet, dann wird dabei stets vorausgesetzt, dass diese Teilmengen einer entsprechend vorgegebenen Grundmenge sind. Andernfalls könnte man zum Beispiel das Komplement einer Menge gar nicht bilden. Ähnlich geht man bei statistischen Untersuchungen in der Regel von der Abgrenzung der sogenannten Grundgesamtheit aus. Es handelt sich dabei um die einer Untersuchung zugrunde liegenden Individuen oder Objekte. Beispiele für Grundgesamtheiten könnten etwa sein

die Wahlberechtigten eines Staates
die Einwohner einer Stadt
die Kunden einer Bank
die Studierenden eines Studiengangs
die PCs eines Unternehmens

Betrachtet man die Gesamtheit der Wahlberechtigten, so erhält man an einem Wahltag natürlich Informationen über deren Beteiligung an der Wahl und über die jeweiligen Parteipräferenzen. Oft besteht aber auch ein weiter reichendes Interesse an der Einstellung der Bevölkerung im Hinblick auf persönliche Ansichten, politische Vorhaben oder Entscheidungen. Dabei wird versucht, mit Hilfe von Umfragen durch Markt- und Meinungsforschungsinstitute, entsprechende Antworten von den Befragten zu erhalten. An dieser Stelle kommt ein wichtiger Begriff ins Spiel, der in gewisser Hinsicht einen Gegensatz zur Grundgesamtheit bildet: nämlich die Stichprobe. Es handelt sich hierbei um eine Auswahl von „Objekten“ einer Grundgesamtheit, wobei der Auswahlprozess häufig zur Gänze oder zumindest teilweise zufällig ist. Zweck einer Stichprobenerhebung ist es, Informationen über verschiedene Aspekte einer Grundgesamtheit zu erhalten, ohne allerdings die Grundgesamtheit als Ganzes untersuchen zu müssen. Letzteres wird zum Beispiel vor allem aus Zeit- oder Kostengründen zweckmäßig sein. Die mit Hilfe einer Stichprobe gewonnene Information ist naturgemäß unvollständig und daher mit einer entsprechenden Unsicherheit behaftet. Auf dieses Problem der Unsicherheit im Zusammenhang mit Stichproben werden wir im Rahmen der Induktiven Statistik zurückkommen.

Natürlich verwendet man auch häufig den Begriff der Grundgesamtheit, ohne dass irgendeine Bezugnahme zu einer Stichprobe besteht. Eine Bestandsaufnahme oder Überprüfung sämtlicher PCs eines Unternehmens wäre dafür ein einfaches Beispiel. Ein weiteres wären die zu Beginn dieses Kapitels beschriebenen Melange-Preise Wiener Kaffeehäuser.

1.2. Merkmale und Verteilungen

Im Hinblick auf die Grundgesamtheit ‘die Einwohner einer Stadt’ könnte man an der Beantwortung von Fragen wie zum Beispiel nach dem Geschlecht, dem Alter, dem Familienstand oder der Stellung im Beruf interessiert sein. Eine Bank wird sich unter anderem für das Einkommen, das Vermögen oder die Altersstruktur ihrer Kunden interessieren. Dies führt uns zum wichtigen Begriffspaar Merkmal (auch Variable genannt) und Merkmalsausprägungen.

Merkmal und Merkmalsausprägungen

Ein Merkmal ist eine Zusammenfassung von Merkmalsausprägungen. Darunter versteht man Zahlenwerte oder Attribute, die den Objekten der Grundgesamtheit zugeordnet werden.

Bei den Melange-Daten handelt es sich somit um eine Untersuchung des Merkmals ‘Preis für eine Melange’, wobei die angegebenen Preise konkret beobachtete Ausprägungen dieses Merkmals darstellen.

Verständlicherweise soll eine eindeutige Zuordnung der Objekte der Grundgesamtheit zu den Merkmalsausprägungen erreicht werden. Dazu ist es erforderlich, dass die Ausprägungen eine Zerlegung (Partition) des Merkmals bilden. Das bedeutet, dass die einzelnen Ausprägungen sich gegenseitig ausschließen und jedem Objekt der Grundgesamtheit genau eine dieser Ausprägungen zugeordnet wird. Es könnte zum Beispiel sein, dass bei einer Umfrage das Merkmal ‘Familienstand’ erhoben wird. Sinnvollerweise sollten dann nicht nur die Ausprägungen ledig, verheiratet, verwitwet bzw. geschieden vorgesehen sein, sondern auch die Möglichkeit, dass die Antwort verweigert wird.

Quantitative und qualitative Merkmale

Quantitative Merkmale sind Merkmale, deren Ausprägungen durch Zahlenwerte beschrieben werden.

Qualitative Merkmale sind Merkmale, deren Ausprägungen durch Attribute beschrieben werden.

Es gibt eine Reihe verschiedener Möglichkeiten, Merkmale einzuteilen. Eine wichtige Unterteilung ist dabei die in quantitative und qualitative Merkmale. Dies hat vor allem Konsequenzen im Hinblick auf die Anwendung spezifischer statistischer Methoden.

Typische Beispiele für quantitative Merkmale sind etwa das Einkommen oder das Lebensalter von Personen, der Kurs einer Aktie oder die Zahl der täglich eintreffenden E-Mails. Die Melange-Preise gehören natürlich ebenfalls zu den quantitativen Merkmalen. Beispiele für qualitative Merkmale sind das Geschlecht, die Stellung im Beruf (Arbeiter, Angestellter, Beamter, Selbständiger), der Familienstand oder der ordentliche Wohnsitz. Quantitative Merkmale lassen sich noch weiter unterteilen in diskrete und stetige Merkmale.

Diskrete und stetige Merkmale

Diskrete Merkmale sind Merkmale, deren Ausprägungen sich der Reihe nach aufzählen lassen.

Stetige Merkmale sind Merkmale, deren Ausprägungen zumindest prinzipiell jeden Wert innerhalb eines bestimmten Intervalls annehmen können.

Typische Beispiele für diskrete Merkmale sind etwa die Anzahl der innerhalb eines bestimmten Zeitraums eintreffenden Ereignisse (zum Beispiel Geburten, Todesfälle), die Kinderzahl pro Familie usw. Zu den stetigen Merkmalen gehören etwa das Alter, die Größe einer Person oder auch die Zeit, die Sie jede Woche vor dem PC verbringen. In diesem Zusammenhang spricht man auch von diskreten und stetigen Daten. Merkmale, die sich in Geldeinheiten ausdrücken lassen (monetäre Größen), werden übrigens häufig wie stetige Merkmale behandelt, obwohl sie ihrer Natur nach eigentlich zu den diskreten Merkmalen gehören. Beispiele dafür sind das Einkommen privater Haushalte, der Umsatz von Unternehmen oder auch die Kurse von Aktien.

Manche statistische Daten kann man an einem bestimmten Stichtag erheben, zum Beispiel die Bevölkerungszahl oder die Zahl der Kunden einer Bank. Solche Daten werden auch Bestandsdaten genannt. Andere Daten lassen sich sinnvoll nur über einen gewissen Zeitraum erheben, wie zum Beispiel Geburten, Todesfälle oder auch das Einkommen. Diese Daten nennt man Bewegungs- oder Stromdaten.

Eine weitere wichtige Klassifizierung von Merkmalen stellt das Skalenniveau dar. Dieses sieht eine Unterteilung von Merkmalen gemäß den folgenden Skalen vor: Nominalskala, Ordinalskala (Rangskala) und Verhältnisskala (metrische Skala). Je nachdem, auf welchem Skalenniveau sich eine Variable befindet, lassen sich bestimmte Operationen durchführen. Qualitative Variablen weisen demnach eine Nominalskala auf. Die Ausprägungen stellen in diesem Fall unterschiedliche Bezeichnungen dar, bei denen aber keine Anordnung wie etwa größer/kleiner vorliegt. Bei Ausprägungen, die eine Rangordnung darstellen, zum Beispiel Ratings von Unternehmen, ist es nicht sehr sinnvoll, diese zu addieren (auch wenn sie in Form von Zahlen vorliegen sollten). Metrische Variablen treten sehr häufig auf, wie zum Beispiel die Preise von Waren oder Dienstleistungen, das Einkommen, die Körpergröße oder das Alter einer Person.

> **Verteilung**
>
> Unter der Verteilung eines Merkmals versteht man die Zuordnung der Merkmalsausprägungen zu den Objekten der Grundgesamtheit.

Der Begriff der Verteilung gehört zu einem der wichtigsten Begriffe der Statistik. Gelegentlich wird die Statistik auch als die Lehre von den Verteilungen bezeichnet. Diese Definition greift vielleicht etwas zu kurz, unterstreicht aber die zentrale Bedeutung dieses Begriffs.

Im Rahmen der Deskriptiven Statistik werden wir nur Verteilungen betrachten, die auf Beobachtungen beruhen, das heißt auf beobachteten Daten. Derartige Verteilungen werden auch als empirische Verteilungen bezeichnet. Von besonderem Interesse ist dabei die Frage: Welche Größen und welche Darstellungsformen verwendet man im Zusammenhang mit Verteilungen? Zunächst zum ersten Teil der Frage. Zur Darstellung (empirischer) Verteilungen verwendet man typischerweise absolute oder relative Häufigkeiten. Haben wir ein qualitatives Merkmal oder

Tabelle 1.2. Die einzelnen Melange-Preise

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

ein diskretes Merkmal, dann versteht man unter der absoluten Häufigkeit einer Ausprägung die Anzahl der Objekte der Grundgesamtheit, die diese Ausprägung besitzen. Für die Melange-Preise zeigt Tabelle 1.2 die Einzelpreise zwischen 2,20 Euro und 4,50 Euro, wobei 10 Cent-Intervalle verwendet wurden, um etwaige Lücken zu vermeiden.

Symbolisch bezeichnet man die absolute Häufigkeit einer Ausprägung i häufig mit h_i. Die relative Häufigkeit f_i erhält man, indem man die absolute Häufigkeit h_i durch die Anzahl n der Objekte der Grundgesamtheit (den Umfang der Grundgesamtheit) dividiert:

$$f_i = \frac{h_i}{n}$$

Die Summe der relativen Häufigkeiten ergibt natürlich den Wert Eins

$$\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} \frac{h_i}{n} = \frac{1}{n} \sum_{i=1}^{k} h_i = 1$$

wobei k die Anzahl der verschiedenen Merkmalsausprägungen bedeutet. In der folgenden Tabelle werden die Melange-Preise noch einmal dargestellt, diesmal in einer etwas übersichtlicheren Form und zwar mit Hilfe von Intervallen bzw. Klassen. Dabei sind jeweils Preise innerhalb eines bestimmten Bereichs zu einer gemeinsamen Klasse zusammengefasst. Mit anderen Worten, die Ausprägungen im Bereich zwischen 2,20 Euro bis unter 2,60 Euro bilden die erste Preisklasse, usw. Die absolute Häufigkeit einer Klasse ist dann die Zahl derjenigen Preise, die genau in diesen Bereich fallen. Ähnlich verhält es sich mit der relativen Häufigkeit einer Klasse.

Tabelle 1.3. Verteilung der Melange-Preise

Preisklasse (von … bis unter)	absolute Häufigkeit	relative Häufigkeit (in %)
2,20 – 2,60	8	11,4
2,60 – 3,00	22	31,4
3,00 – 3,40	22	31,4
3,40 – 3,80	12	17,1
3,80 – 4,20	1	1,4
4,20 – 4,60	5	7,1
Summe	70	100,0

Eine derartige Darstellung, bei der Merkmalsausprägungen in Klassen zusammengefasst werden und die absoluten und relativen Häufigkeiten auf der Grundlage dieser Klassen gebildet werden, ist bei stetigen Merkmalen allgemein üblich. Eine Altersverteilung könnte zum Beispiel auf der Klasseneinteilung 0–9 Jahre, 10–19 Jahre, 20–29 Jahre usw. beruhen. Absolute und relative Häufigkeiten beziehen sich dann immer auf die jeweiligen Klassen, wobei es dabei auch zu ungleichen Klassenbreiten kommen kann (Einkommensverteilung).

Gelegentlich werden in Tabellen die kumulierten absoluten und relativen Häufigkeiten ausgewiesen. Die folgende Tabelle zeigt dies für die Melange-Preise. Dabei enthält die zweite Spalte die Summe der kumulierten absoluten Häufigkeiten (hier mit H_i bezeichnet), während die dritte Spalte die entsprechenden kumulierten relativen Häufigkeiten enthält (hier mit F_i bezeichnet). Dieser Tabelle kann man zum Beispiel entnehmen, dass 52 von insgesamt 70 Preisen, das heißt etwa 75 %, unter 3,40 Euro liegen. Entsprechend liegen dann etwa 25 % der Preise bei 3,40 Euro oder darüber.

TABELLE 1.4. Verteilung der Melange-Preise

Preisklasse (von … bis unter)	H_i	F_i (in %)
2,20 – 2,60	8	11,4
2,60 – 3,00	30	42,9
3,00 – 3,40	52	74,3
3,40 – 3,80	64	91,4
3,80 – 4,20	65	92,9
4,20 – 4,60	70	100,0

1.3. Tabellen und Grafiken

Im Folgenden werden wir etwas näher auf verschiedene Darstellungsformen statistischer Daten eingehen. Einige Beispiele dazu haben wir ja bereits kennengelernt. Nicht wenige assoziieren Statistik gerade mit der Vorstellung von Tabellen. Dabei muss man nicht einmal an Wirtschaftsdaten denken. Viele Kinder und Jugendliche dürften mit verschiedenen Tabellen aus dem Bereich des Sports vermutlich recht vertraut sein. Man denke nur an die in vielen Medien gegenwärtigen Fußballtabellen[4] oder die Medaillenspiegel bei Olympischen Spielen, Welt- und Europameisterschaften. Vielleicht könnte man hier in gewisser Hinsicht von einer Art Urerfahrung der Statistik sprechen.

Tabellen liefern in anschaulicher Form Informationen über die jeweilige Verteilung, was man insbesondere auch für Vergleichszwecke verwenden kann. Da für die Melange-Preise zwei Teil-Datensätze zur Verfügung stehen, kann man diese etwa im Rahmen einer größeren Tabelle zusammenfassen und einander gegenüberstellen (siehe Tabelle 1.5). Vorteilhaft ist dabei natürlich die Verwendung gleicher Preisklassen. Als Träger detaillierter Informationen über Daten eignen sich Tabellen auch besonders als Quellenangaben.

Tabelle 1.5. Verteilung der Melange-Preise: 1. Bezirk und alle übrigen Bezirke

	1. Bezirk		alle übrigen Bezirke	
Preisklasse (von ... bis unter)	absolute Häufigkeit	relative Häufigkeit (in %)	absolute Häufigkeit	relative Häufigkeit (in %)
2,20 – 2,60	1	3,3	7	17,5
2,60 – 3,00	3	10,0	19	47,5
3,00 – 3,40	11	36,7	11	27,5
3,40 – 3,80	10	33,3	2	5,0
3,80 – 4,20	1	3,3	0	0,0
4,20 – 4,60	4	13,3	1	2,5
Summe	30	100,0	40	100,0

Gegenüber Tabellen[5] verlieren grafische Darstellungen in der Regel an Präzision, liefern dafür aber einen besseren Überblick über die Form der Verteilung. Zur Anwendung kommen dabei häufig Histogramme, Stabdiagramme oder Kreisdiagramme, von denen wir später noch einige Beispiele sehen werden. Zu beachten ist dabei, dass die geometrische Form einer Darstellung mit der Häufigkeit konform gehen sollte. Verwendet man etwa ein Histogramm zur Darstellung einer Häufigkeitsverteilung, so werden in diesem Fall natürlich die Rechteckflächen des Histogramms als Häufigkeiten interpretiert (nicht die Höhen). Ähnliches gilt entsprechend auch für die Darstellung von Kreisdiagrammen. Bei den Abbildungen 1.1 und 1.2 werden als Beispiele Histogramme zur Darstellung der Melange-Preise verwendet.

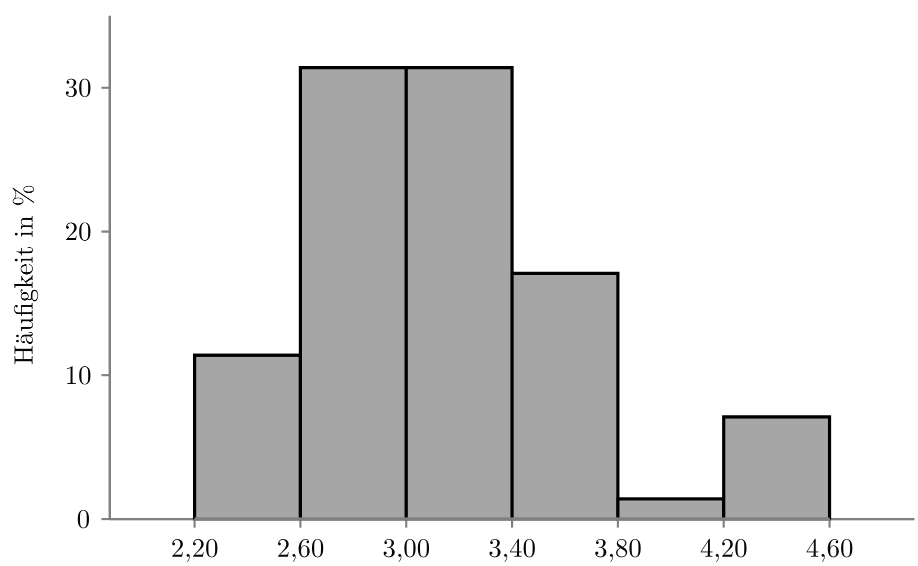

Abbildung 1.1. Histogramm der Verteilung der Melange-Preise

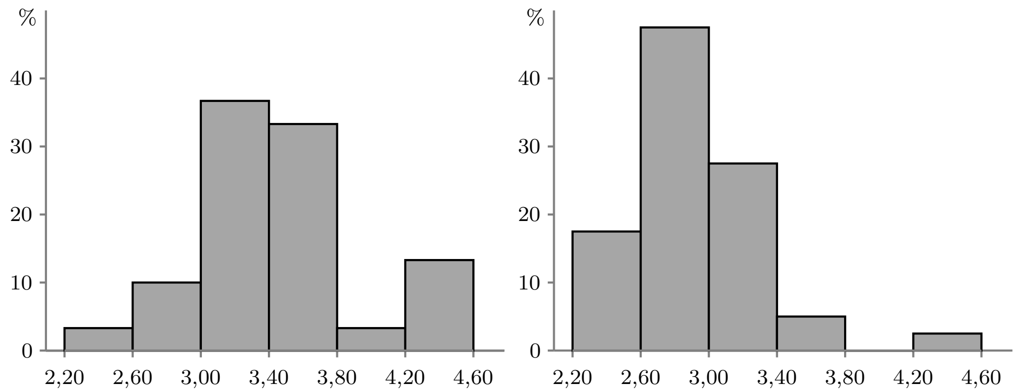

Abbildung 1.2. Verteilung der Melange-Preise: 1. Bezirk (linke Grafik) und alle übrigen Bezirke (rechte Grafik)

Eine interessante Kombination aus Tabelle und Grafik stellen die sogenannten Stiel-und-Blatt-Diagramme dar. Ein Beispiel dazu ist der nachfolgende Fahrplan einer Straßenbahnlinie[6]. Durch die spezielle Anordnung der Abfahrtszeiten (immerhin etwa 450 Werte) liefert diese Darstellung nicht nur die präzisen Informationen einer Tabelle, sondern gleichzeitig einen grafikähnlichen Überblick über die stündlichen Abfahrtshäufigkeiten. Ganz nebenbei erhält man hier auch noch einen übersichtlichen Vergleich verschiedener Einzelfahrpläne.

Tabelle 1.6. Straßenbahnfahrplan in Form von Stiel-und-Blatt-Diagrammen

Montag–Freitag (Schule)	Montag–Freitag (Ferien)	Samstag	Sonn- und Feiertag
5 31 41 51	**5** 31 41 51 59	**5** 21 31 41 53	**5** 24 41 57
6 00 08 15 23 30 38 45 53	**6** 08 15 23 30 38 45 53	**6** 03 13 23 33 43 53	**6** 12 27 42 57
7 00 08 15 23 30 37 43 49 55	**7** 00 08 15 23 30 38 45 53	**7** 03 13 23 33 43 53	**7** 12 27 42 57
8 01 08 15 23 30 38 45 53	**8** 00 08 15 23 33 43 53	**8** 03 13 23 33 43 53	**8** 12 23 33 43 53
9 00 08 15 23 30 38 45 53	**9** 03 13 23 33 43 53	**9** 03 13 23 33 43 53	**9** 03 13 23 33 43 53
10 00 08 15 23 30 38 45 53	**10** 03 13 23 33 43 53	**10** 03 13 23 33 43 53	**10** 03 13 23 33 43 53
11 00 08 15 23 30 38 45 53	**11** 03 13 23 33 43 53	**11** 03 13 23 33 43 53	**11** 03 13 23 33 43 53
12 00 08 15 23 30 38 45 53	**12** 03 13 23 33 43 53	**12** 03 13 23 33 43 53	**12** 03 13 23 33 43 53
13 00 08 15 23 30 38 45 53	**13** 03 13 23 33 43 53	**13** 03 13 23 33 43 53	**13** 03 13 23 33 43 53
14 00 08 15 23 30 38 45 53	**14** 03 13 23 30 38 45 53	**14** 03 13 23 33 43 53	**14** 03 13 23 33 43 53
15 00 08 15 23 30 38 45 53	**15** 00 08 15 23 30 38 45 53	**15** 03 13 23 33 43 53	**15** 03 13 23 33 43 53
16 00 08 15 23 30 38 45 53	**16** 00 08 15 23 30 38 45 53	**16** 03 13 23 33 43 53	**16** 03 13 23 33 43 53
17 00 08 15 23 30 38 45 53	**17** 00 08 15 23 30 38 45 53	**17** 03 13 23 33 43 53	**17** 03 13 23 33 43 53
18 00 08 15 23 30 38 45 53	**18** 00 08 15 23 30 38 45 53	**18** 03 13 23 33 43 53	**18** 03 13 23 33 43 53
19 00 08 15 23 33 43 53	**19** 00 08 15 23 33 43 53	**19** 03 13 23 33 43 53	**19** 03 13 23 33 43 53
20 03 13 23 33 43 53	**20** 03 13 23 33 43 53	**20** 03 13 27 42 57	**20** 03 13 27 42 57
21 03 13 23 33 43 57	**21** 03 13 23 33 43 57	**21** 12 27 42 57	**21** 12 27 42 57
22 12 27 42 57	**22** 12 27 42 57	**22** 12 27 42 57	**22** 12 27 42 57
23 12 27 42 57	**23** 12 27 42 57	**23** 12 27 42 57	**23** 12 27 42 57
0 12 27 39	**0** 12 27 39	**0** 12 27 39	**0** 12 27 39
1 00	**1** 00	**1** 00	**1** 00

Die spezielle Bezeichnung dieser Diagramme wird verständlich, wenn man sich die Spalten der Stunden (von 5 bis 1) als Stiel und die dazugehörigen Abfahrtszeiten als entsprechende Blätter einer Blume vorstellt.

In den folgenden Kapiteln werden wir es häufig mit Zeitreihen zu tun haben. Bei einer Zeitreihe handelt es sich um zeitlich angeordnete Daten. Bei Wirtschaftsdaten hat man es dabei vor allem mit jährlichen, vierteljährlichen oder monatlichen Daten zu tun, gelegentlich auch mit Wochen- oder Tagesdaten. Beispiele dafür sind die Entwicklung des privaten Konsums, die jährlichen Umsatzzahlen eines Unternehmens oder die Darstellung verschiedener Finanzdaten.

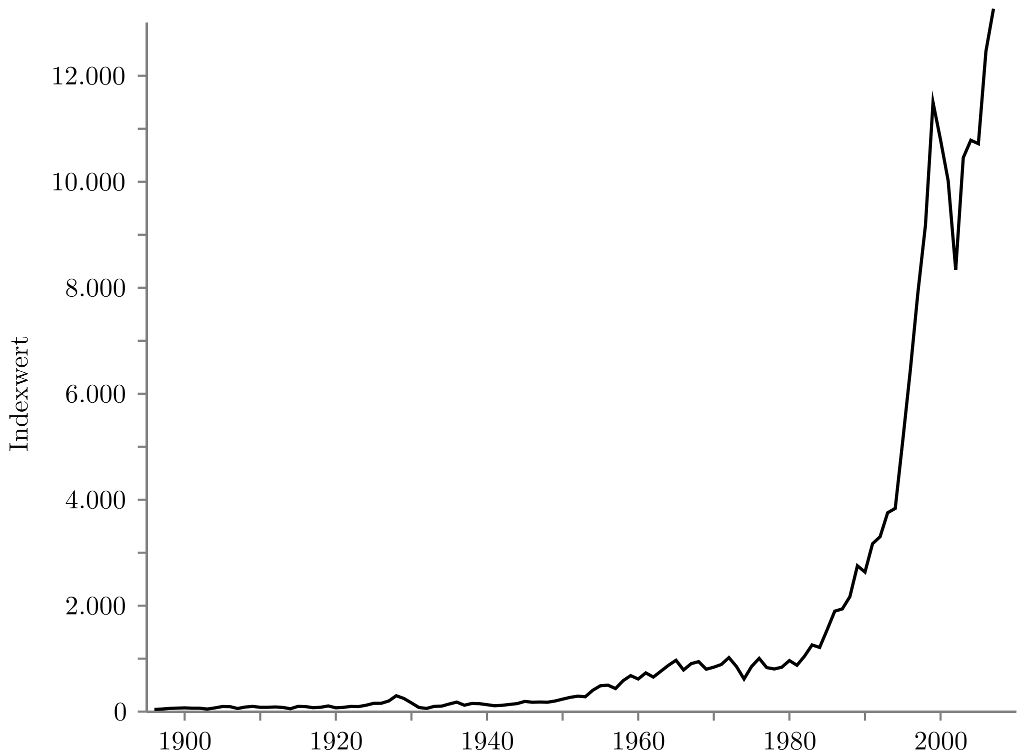

Abbildung 1.3. Dow-Jones-Index 1896–2007 (Jahresendwerte)

Betrachten wir dazu als Beispiel eine Zeitreihe des US-amerikanischen Dow-Jones-Index, des international bekanntesten Aktienindex[7]. Er wurde erstmals im Jahre 1896 für eine Auswahl von Aktien der New Yorker Börse berechnet und setzt sich aus insgesamt 30 Aktientiteln zusammen. Abbildung 1.3 zeigt den Verlauf des Index über den Zeitraum 1896–2007 auf der Basis der jeweiligen Jahresendwerte. Für jedes Jahr des genannten Zeitraums ist der Schlusswert[8] des letzten Handelstages angegeben. An diesem Beispiel kann man insbesondere die Schwierigkeiten erkennen, die sich aus dem gleichzeitigen Auftreten von kleinen und großen Werten innerhalb einer Grafik ergeben können.

Die Darstellung zeigt bis Ende der 1940er Jahre einen geringfügigen Anstieg des Index, der bis Ende der 1970er Jahre etwas stärker ausgeprägt ist und erst in den 1980er Jahren zu einer extremen Dynamik ansetzt. Wer etwas mit der Wirtschafts- bzw. Finanzgeschichte vertraut ist, wird mit einem gewissen Erstaunen feststellen, dass etwa der Börsenkrach des Jahres 1929 und die sich daraus entwickelnde Weltwirtschaftskrise ("Great Depression") der 1930er Jahre oder auch die erste Erdölkrise der Jahre 1973/74 nur als winzige Details in der Grafik zu erkennen sind. Diese Einbrüche hatte man eigentlich anders in Erinnerung. Das hat natürlich mit den hohen Indexwerten der 1990er Jahre zu tun. Ein Rückgang des Index zum Beispiel um 30 Punkte bedeutete Ende der 1920er Jahre einen zweistelligen Renditeverlust, während dies bei einem Kursniveau in der Höhe von 10.000 Punkten praktisch nicht wahrgenommen wird.

Mit anderen Worten, die historische Entwicklung des Dow-Jones-Index wird anhand dieser Grafik äußerst unzureichend wiedergegeben. Der Grund dafür liegt auf der Hand. Investoren orientieren sich nicht so sehr an absoluten Kursen, sondern vor allem an Renditen, und deren Entwicklung wird von der Grafik nicht erfasst! Dazu gibt es allerdings eine Alternative. Die Darstellung lässt sich nämlich wesentlich zweckmäßiger gestalten, wenn man zu einer logarithmischen Skalierung übergeht, bei der statt der absoluten die logarithmierten Indexwerte verwendet werden. Diese Vorgangsweise lässt sich wie folgt begründen. Entsprechend der klassischen Renditeformel besitzt die Rendite r des Index über die Zeitperiode von t nach $t+1$ den Wert

$$r = \frac{x_{t+1} - x_t}{x_t} = \frac{x_{t+1}}{x_t} - 1$$

wobei x_t und x_{t+1} die jeweiligen Indexwerte am Beginn und am Ende der Zeitperiode bedeuten. Für den Logarithmus von $1+r$ folgt dann

$$\ln(1+r) = \ln\left(\frac{x_{t+1}}{x_t}\right) = \ln(x_{t+1}) - \ln(x_t)$$

wobei bei der zweiten Gleichung eine Rechenregel für Logarithmen verwendet wurde. Man kann zeigen, dass die Approximation $\ln(1+r) \approx r$ für moderate Werte von r eine recht gute Näherung darstellt[9].

Es gilt somit die Approximation $r \approx \ln(x_{t+1}) - \ln(x_t)$ für nicht zu große Werte von r. Dies bedeutet aber nichts anderes, als dass der Unterschied zwischen zwei logarithmierten Indexwerten in etwa der Rendite über die jeweilige Zeitperiode entspricht! Anders ausgedrückt: durch die Logarithmierung der Indexwerte wird gerade die Renditeentwicklung des Index sichtbar. Genau das ist es aber, was für die Finanzwelt von Interesse ist[10].

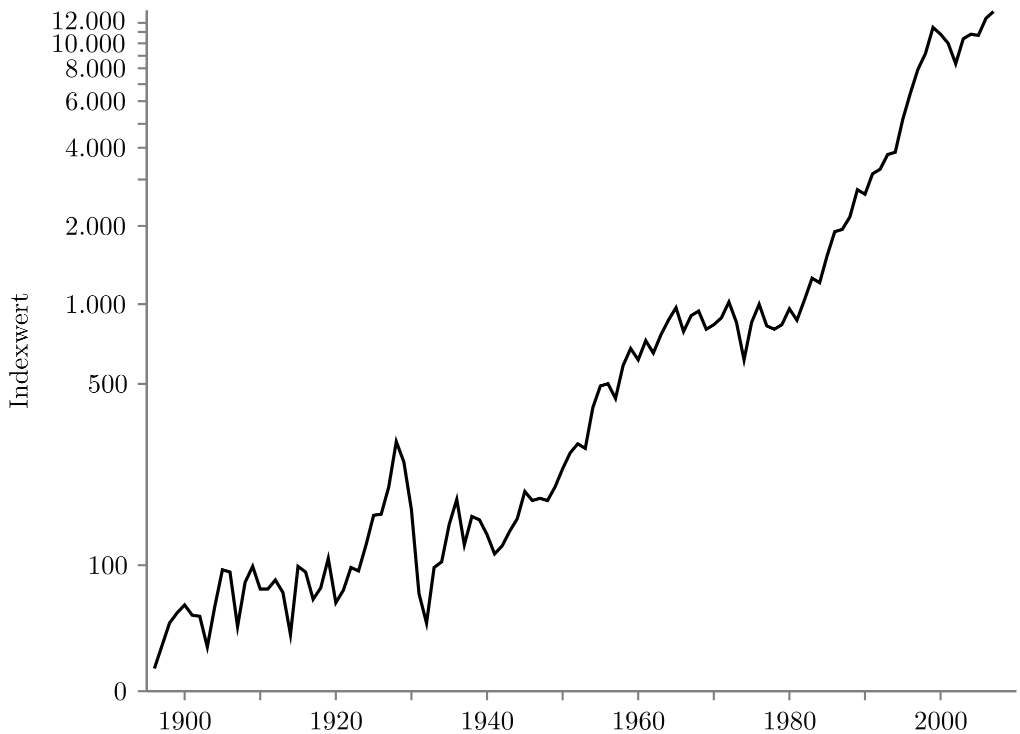

Abbildung 1.4. Dow-Jones-Index 1896–2007 (Jahresendwerte, logarithmiert)

Die obige Grafik zeigt die logarithmierten Indexwerte, wodurch die historische Entwicklung des Index wesentlich transparenter und damit auch realistischer wiedergegeben wird. Diese Grafik erzählt jetzt eine ganz andere Geschichte. Die größte Änderung erkennt man genau dort, wo man sie aus historischen Gründen auch erwarten würde, nämlich gegen Ende der 1920er und zu Beginn der 1930er Jahre. Größere (insbesondere negative) Änderungen in den Renditen sind vor allem in den ersten Jahrzehnten des 20. Jahrhunderts zu beobachten. Ab den 1940er Jahren werden die Ausschläge (nach unten) deutlich geringer, insbesondere ab den 1980er Jahren.

1.4. Aktienmärkte. Eine empirische Fallstudie (1)

In diesem und einigen nachfolgenden Kapiteln werden wir uns im Rahmen einer kleinen empirischen Fallstudie etwas näher mit Aktienindizes beschäftigen. Unter einem Aktienindex versteht man ganz allgemein ein Portfolio von speziell ausgewählten Aktien eines Marktes, wobei dieses Portfolio den Markt repräsentieren und dessen durchschnittliche Kursentwicklung beschreiben soll. Am Beispiel von vier Indizes werden wir verschiedene statistische Aspekte des jeweiligen