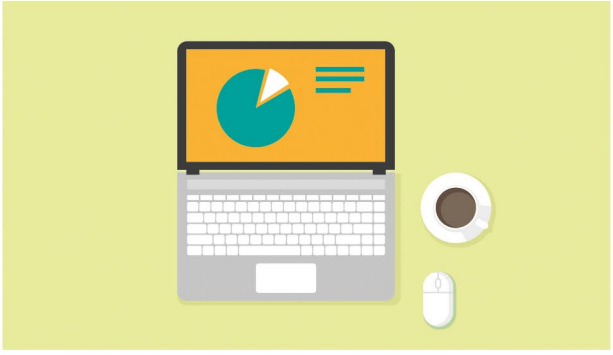
2. Auflage

# Fit fürs Studium

# Statistik

+++ Benno Grabinger +++



+++ Grundkonzepte verstehen und Wissenslücken schließen. Ideal zum Selbststudium. Mit vielen realen Beispielen, Aufgaben und ausführlichen Lösungen. +++

Mit Übungsmaterial zum Download

Rheinwerk Computing

# Liebe Leserin, lieber Leser,

aus Gesprächen mit Lehrenden und Studierenden wissen wir: Jedes Studium setzt Inhalte aus dem »Schulstoff« voraus. Hat die Vorlesung einmal begonnen und ist der Takt vorgegeben, sollten diese Inhalte präsent sein, sonst wissen sich Hörer wie Dozenten nur schwer zu helfen.

Also heißt es: Wissenslücken vorab schließen. Damit das mit Schwung und ohne aufwändige Recherche gelingt, haben wir »Fit fürs Studium« im Programm. Es spielt keine Rolle, für welchen Studiengang oder für welchen Beruf Sie Statistik brauchen: Unser Autor führt Sie mit vielen anschaulichen Beispielen in alle wichtigen Themen ein. Die Daten und Fragestellungen stammen aus allgemeinverständlichen Bereichen – Sie rechnen mit Kirmeslosen, Gesundheitsangaben, Längen von Gegenständen oder Würfeln. Jedes Kapitel steigt mit einem konkreten Beispiel ein, das intuitiv zugänglich ist. Mit den Aufgaben lernen Sie nach und nach, die Theorie anzuwenden, mit der formalen Sprache umzugehen und immer mehr Berechnungen selbst durchzuführen.

Idealerweise arbeiten Sie das Buch von vorne nach hinten durch, denn die meisten Inhalte bauen aufeinander auf, wie es für die Mathematik typisch ist. Mit passenden Vorkenntnissen lassen sich die Kapitel aber auch einzeln lesen. Das bietet sich an, falls Sie nur an einem bestimmten Thema interessiert sind oder später noch einmal etwas nachlesen möchten. Wenn dann doch eine Voraussetzung nicht präsent sein sollte, helfen Ihnen die Querverweise auf nummerierte Abschnitte, Beispiele und Aufgaben.

Damit Sie mit Daten und Zusammenhängen experimentieren können, ohne viel tippen zu müssen, gibt es zu vielen Abschnitten Übungsmaterial zum Herunterladen. Scrollen Sie dazu auf der Seite zum Buch unter *https://www.rheinwerk-verlag.de/5388* etwas herunter bis zu den MATERIALIEN ZUM BUCH. Die Schaltfläche ZU DEN MATERIALIEN führt Sie zum Download.

Eine Anmerkung noch in eigener Sache: Wir möchten unsere Arbeit und unsere Bücher immer besser machen. Feedback und konstruktive Kritik sind uns deshalb sehr willkommen!

**Ihre Almut Poll**
Lektorat Rheinwerk Computing

almut.poll@rheinwerk-verlag.de
www.rheinwerk-verlag.de
Rheinwerk Verlag · Rheinwerkallee 4 · 53227 Bonn

# Auf einen Blick

# Impressum

Dieses E-Book ist ein Verlagsprodukt, an dem viele mitgewirkt haben, insbesondere:

**Lektorat** Almut Poll
**Fachgutachten** Johannes Schneider
**Korrektorat** Joram Seewi
**Herstellung E-Book** Melanie Zinsler, Norbert Englert
**Covergestaltung** Mai Loan Nguyen Duy
**Coverbilder** unsplash © Joanna Kosinska; iStock: 499629512 © Eva-Katalin, 509290786 © vgajic, 139887145 © Gloda
**Satz E-Book** SatzPro, Krefeld

Wir hoffen sehr, dass Ihnen dieses Buch gefallen hat. Bitte teilen Sie uns doch Ihre Meinung mit und lesen Sie weiter auf den *Serviceseiten*.

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

**ISBN 978-3-8362-8618-3 (E-Book)**
**ISBN 978-3-8362-8611-4 (Bundle)**

2., korrigierte Auflage 2021


www.rheinwerk-verlag.de

*Für Sara und Elisa*

# Inhalt

## TEIL III Beurteilende Statistik

# Über dieses Buch

### Die Situation

Löst das Wort »Statistik« bei Ihnen Panik aus, besonders dann, wenn Sie feststellen müssen, dass Statistik ein Bestandteil Ihres Studiums ist und Ihnen die Grundlagen aus der Schulzeit fehlen?

Neben Mathematik, Wirtschaftswissenschaften und Naturwissenschaften sind auch Psychologie, Soziologie, Medizin, Pädagogik und Geowissenschaften typische Beispiele für Studiengänge, in denen ein Statistikkurs Bestandteil der Studienordnung ist.

Können Sie mit Begriffen wie Zufallsvariable, Wahrscheinlichkeitsverteilung, Korrelation, Punktschätzung und Hypothesentests wenig oder gar nichts verbinden, und wissen Sie aber, dass diese Begriffe Grundvoraussetzung für eine erfolgreiche Bewältigung eines Statistikkurses sind? Dann sollten Sie nicht gleich einen Studienfachwechsel oder gar den Abbruch des Studiums in Erwägung ziehen. Schöpfen Sie lieber Mut aus den Worten von René Descartes (1596–1650, Philosoph, Mathematiker und Naturwissenschaftler), der sagte: »Ich glaube, dass selbst zur Entdeckung der schwierigsten Wahrheiten, wenn man nur richtig geleitet wird, nichts als der gesunde Menschenverstand erforderlich ist.« Den gesunden Menschenverstand besitzen Sie, weil Sie sich die Zulassung für ein Studium erarbeitet haben. Nach Descartes fehlt dann nur noch die richtige Anleitung. Dazu macht Ihnen dieses Buch mit seinen vielen anschaulichen Beispielen und dem niedrigen Einstiegsniveau ein Angebot. Wenn Sie weiterlesen, erfahren Sie, wie dieses aussieht.

### Was Sie erwartet

Dieses Buch soll Ihnen die Grundbegriffe von Statistik und Wahrscheinlichkeitsrechnung nahebringen, die erforderlich sind, damit Sie sich erfolgreich auf einen Statistikkurs vorbereiten können. Um jegliches Missverständnis zu vermeiden, muss deutlich gesagt werden, dass dieses Buch Ihren Statistikkurs nicht vorwegnimmt, sondern Ihnen nur die Voraussetzung liefert, diesen erfolgreich zu absolvieren.

Das Buch gliedert sich in drei Hauptteile. Im ersten Teil, der »deskriptiven Statistik«, erfahren Sie, wie man erhobene Daten durch geeignete Kennzahlen wie z. B. Mittelwerte verdichtet und wie man diese Daten dann übersichtlich darstellt. Die Inhalte des zweiten Teils, »Wahrscheinlichkeitsrechnung«, fungieren als Bindeglied, da im dritten Teil Wahrscheinlichkeitsrechnung erforderlich ist. Im Wesentlichen stellt dieser zweite Teil die Inhalte dar, welche Sie aus dem Schulunterricht kennen sollten, sodass Sie sich gezielt mit den Abschnitten beschäftigen können, die Ihnen weniger vertraut sind. Der

dritte Teil, die »beurteilende Statistik«, soll Ihnen die Grundlagen des Schätzens und Testens an grundlegenden Beispielen verständlich machen.

### Wie das Buch aufgebaut ist

Für alle in diesem Buch behandelten Themen gibt es eine gemeinsame Vorgehensweise. Als Ausgangspunkt wird stets ein Beispiel, oft mit realen Daten, gewählt. An diesem Beispiel werden die wesentlichen das Problem betreffenden Punkte herausgearbeitet. Es erfolgt dann eine Verallgemeinerung der erkannten Strukturen. Dabei wird oft, dem Charakter dieses Buches angemessen, auf eine exakte mathematische Behandlung verzichtet und dafür mehr Wert auf einen anschaulichen Zugang gelegt. Es schließen sich Beispiele und Aufgaben an, die das Problem an einer ähnlichen Situation und unter einem leicht veränderten Blickwinkel verdeutlichen. Die Aufgaben sollen Ihnen dazu dienen, zu überprüfen, ob Sie die Problematik erfasst haben und ob Sie die erarbeiteten Begriffe selbst anwenden können. Sie finden die Lösungen der Aufgaben jeweils am Ende des Kapitels. Das erfordert von Ihnen allerdings die Disziplin, nicht gleich die Lösung zu lesen, sondern sich an der Aufgabe erst selbst zu versuchen. Diese Selbsttätigkeit ist für das Lernen von Mathematik stets erforderlich.

Bei komplexeren Rechnungen erhalten Sie dadurch Hilfe, dass Ihnen unter dem Hinweis »*So berechnen Sie…*« die Teile der Rechnung in leicht durchführbaren Einzelschritten präsentiert werden. An dieses Schema können Sie sich halten, wenn Sie unsicher sind.

Am Ende eines jeden Abschnittes finden Sie die Rubrik »*Was Sie wissen sollten*«. Darin sind die wichtigsten Lernziele des letzten Abschnittes aufgeführt, und Sie können damit überprüfen, wie fit Sie schon sind.

### Welche Voraussetzungen müssen Sie mitbringen?

Mit der Bruchrechnung sollten Sie sich auskennen, und auch Prozente sollten Ihnen nicht unbekannt sein. Aus dem Bereich der sogenannten Analysis sollte Ihnen die Euler'sche Zahl $e$ bekannt sein, besser noch die darauf beruhende natürliche Exponentialfunktion. Aus der Integralrechnung reicht es, wenn Sie ein bestimmtes Integral als Flächeninhalt interpretieren können, ohne dass Sie selbst Integrale berechnen müssen. Womit Sie sich auf jeden Fall noch befassen müssen, ist das »Summenzeichen«, das Sie aus dem Schulunterricht wahrscheinlich nicht kennen. Sie benötigen es zum Lesen dieses Buches und für Ihren Statistikkurs. Um Ihnen dabei zu helfen, sich mit diesem Zeichen vertraut zu machen, beschäftigt sich ein eigener Abschnitt des Buches mit dem Summenzeichen und liefert Ihnen eine Reihe von Beispielen und Aufgaben.

### Excel-Dateien

Zu diesem Buch gehören 33 Excel-Tabellenkalkulationsblätter, die Sie auch mit OpenOffice lesen können sollten. Mit diesen Dateien können Sie Simulationen und Beispiele, die im Buch beschrieben sind, nachvollziehen und damit besser verstehen. Im Text des Buches wird jeweils darauf hingewiesen, wann Sie ein Tabellenblatt sinnvoll einsetzen können. Sie sind in Arbeitsmappen mit den Dateinamen *1.xlsx*, ..., *8.xlsx* zusammengefasst. Dabei beziehen sich die Nummern auf die Kapitel im Buch. Außerdem gibt es die Mappe *Tabellen.xlsx*, welche interaktive Tabellen zu folgenden Verteilungen enthält: Binomialverteilung, Poisson-Verteilung, Hypergeometrische Verteilung, Geometrische Verteilung, Normalverteilung und Wilcoxon-Verteilung. Mit diesen Tabellen können Sie leicht Wahrscheinlichkeiten zu den einzelnen Verteilungen berechnen lassen, deshalb beschränkt sich der Tabellenanhang des Buches auf die Standardnormalverteilung. Sämtliche Dateien enthalten Hinweise zur korrekten Verwendung. Außerdem sind alle Dateien schreibgeschützt, sodass auch ein ungeübter Nutzer nichts »zerstören« kann. Weil der Schreibschutz nicht passwortgeschützt ist, können Experten diesen auch entfernen und die Datei ihren Bedürfnissen anpassen. Die Dateien helfen Ihnen sicher beim Studium des Buches und bei eigenen Rechnungen, trotzdem wird darauf hingewiesen, dass es sich nicht um professionell erstellte Dateien handelt, sondern um Dateien, die parallel zum Schreiben dieses Buches entstanden sind.

### Danksagung

Für den familiären Rückhalt während meiner Schreibphase bedanke ich mich bei meiner Ehefrau Hildegard, die in dieser Zeit manches ertragen musste. Bei den Mitarbeitern des Rheinwerk Verlags, allen voran bei meinen Lektorinnen Almut Poll und Anne Scheibe, bedanke ich mich für die vielfältige Unterstützung.

TEIL I

# Deskriptive Statistik

Kapitel 1

# Grundbegriffe der Statistik

*Wir werfen einen Blick auf die Anfänge der Statistik und klären wichtige Begriffe. Erste Beispiele werden diskutiert und erläutert. Außerdem bekommen Sie einen Überblick darüber, wie sich das Gebiet der Statistik heute gliedert und welche Teilgebiete es umfasst.*

## 1.1 Die Anfänge

Zum Einstieg

*Es begab sich aber zu der Zeit, dass ein Gebot von dem Kaiser Augustus ausging, dass alle Welt geschätzt würde. Und diese Schätzung war die allererste und geschah zu der Zeit, da Cyrenius Landpfleger von Syrien war. Und jedermann ging, dass er sich schätzen ließe, ein jeglicher in seine Stadt.*

*Da machte sich auch auf Joseph aus Galiläa, aus der Stadt Nazareth, in das jüdische Land zur Stadt Davids, die da heißt Bethlehem, darum dass er von dem Hause und Geschlechte Davids war, auf dass er sich schätzen ließe mit Maria, seinem vertrauten Weibe, die ward schwanger.*

Im neuen Testament findet man diesen Text bei Lukas 2, 1–5. Es handelt sich um die Erwähnung einer vom römischen Kaiser Augustus befohlenen Volkszählung. Diese wurde von Publius Sulpicius Quirinius (Cyrenius), dem wichtigsten Statthalter Roms im Osten, angeordnet. Jeder musste sich an seinem Herkunftsort in die Steuerlisten eintragen lassen. Nach dem Bericht von Lukas war es dieser Befehl, der Josef und Maria zur Reise nach Bethlehem führte.

Volkszählungen (census) gab es seit dem 5. Jahrhundert v. Chr. im Römischen Reich regelmäßig. Sie dienten unter anderem zur Vermögensschätzung der Bürger und als Unterlage zur Musterung für das Heer. Der Begriff *Zensus* für Totalerhebungen wurde in unsere heutige Statistik übernommen. Statistische Erhebungen von Bevölkerungszahlen sind auch aus China ca. 2000 v. Chr. bekannt. In Ägypten wurden ebenfalls schon um 2700 v. Chr. Bevölkerungszahlen für Steuerzwecke erhoben. Davon zeugen Zensus-Papyri aus pharaonischer Zeit. So gibt es einen Papyrus, heute im Besitz des Britischen Museums London, der aus Theben stammt und eine Aufstellung von Häusern und deren Haushaltsvorstehern enthält [KRS].

Von Inschriften ist bekannt, dass die Ägypter ausführliche Statistiken über die Flut des Nils führten. Jedes Jahr im Juli stieg der Nil an. Erreichte er eine bestimmte Höhe, dann wurden die Dämme durchstochen, sodass sich das Wasser mit dem fruchtbaren Schlamm auf den Feldern verteilte. Nach Ablauf des Wassers konnten die Felder neu eingesät werden. Nach diesen jährlichen Überschwemmungen mussten auf Grundlage der alten Besitzverhältnisse eine neue gerechte Vermessung und Verteilung der Felder erfolgen.

**Abbildung 1.1** Fruchtbares Niltal

Im Mittelalter gab es nur vereinzelte Ansätze zu Volkszählungen. Sie dienten hauptsächlich der Erfassung wehrfähiger Männer. Im Deutschen Kaiserreich wurden dann ab 1871 im Fünfjahreszyklus Volkszählungen durchgeführt.

Die letzte in Deutschland durchgeführte Volkszählung fand im Jahr 2011 statt. Jährlich wird von den Statistischen Landesämtern der sogenannte *Mikrozensus* durchgeführt. Für diese statistische Erhebung wird 1 % der Privathaushalte in Deutschland ausgewählt. Diese Haushalte werden stellvertretend für die gesamte Bevölkerung befragt. Die Auswahl dieser Haushalte geschieht nicht willkürlich, sondern nach statistischen Regeln, um zu gewährleisten, dass eine sogenannte *repräsentative Stichprobe* vorliegt. Deshalb kann auch kein Ausgewählter von der Auskunftspflicht befreit werden, auch nicht, wenn alters- oder krankheitsbedingte Umstände vorliegen. Damit ist gewährleistet, dass alle Bevölkerungsgruppen in der Stichprobe gleichmäßig vertreten sind. Die erhobenen Daten bilden die Grundlage für viele Entscheidungen im Bereich des öffentlichen Lebens. Der Bedarf an Kindergartenplätzen, Schulen und Krankenhäusern wird ebenso daraus geschätzt wie die Nachfrage an Wohnraum.

Am Beispiel des Mikrozensus lassen sich die Aufgaben der Statistik erkennen: Es werden **Daten aus der Wirklichkeit erfasst und durch geeignete Kennzahlen wie z. B. Mittelwerte verdichtet**. Mit Diagrammen werden diese Daten dann übersichtlich dargestellt. Im Groben ist das die Aufgabe der *beschreibenden (deskriptiven) Statistik*. Was sich aus den Diagrammen augenscheinlich aufdrängt, z. B. Unterschiede zwischen ver-

schiedenen untersuchten Gruppen, führt dann zu Hypothesen, die getestet werden müssen. Dazu sind Methoden der Wahrscheinlichkeitsrechnung erforderlich. Dieses Teilgebiet der Statistik heißt *schließende* oder *induktive Statistik*.

### *Deskriptive und induktive Statistik*

**Deskriptive Statistik** ist die **Erhebung und Betrachtung** von Daten.

**Induktive Statistik** bedeutet, dass aus den erhobenen Daten **Schlüsse gezogen** werden. Zum Beispiel sollen aus den Daten Schätzwerte abgeleitet werden, oder es sollen anhand der Daten **Hypothesen** getestet werden.

Das folgende Beispiel zeigt eine typische Aufgabenstellung der beschreibenden Statistik.

**Aufgabe 1: Drogeriemarkt-Umsätze**

Eine Drogeriemarktkette besteht aus 17 Filialen. Die jährlichen Umsätze der einzelnen Filialen in der Einheit 100 000 € sind: 5,25; 6,15; 5,67; 1,365; 4,2; 6,84; 1,095; 4,725; 3,6; 4,8; 2,7; 2,64; 1,02; 1,47; 2,01; 4,425; 3,1.

Zur besseren Übersichtlichkeit sollen verschiedene Parameter gebildet werden.

a) Gesamtumsatz aller Filialen
b) Der mittlere Umsatz (arithmetisches Mittel)
c) Der Umsatz der drei kleinsten und der Umsatz der drei größten Filialen
d) Für die Klasseneinteilung [0; 1), [1; 2), [2; 3), [3; 4), [4; 5), [5; 6), [6; 7), ebenfalls in der Einheit 100 000 €, soll ermittelt werden, wie viele Filialen zu jeder der Klassen gehören. Bei der Schreibweise soll die eckige Klammer andeuten, dass der danebenstehende Wert zur Klasse gehört, die runde Klammer hingegen besagt, dass der danebenstehende Wert *nicht* zur Klasse gehört.

Gehen Sie diese Fragestellungen nun der Reihe nach an.

Schauen Sie am Ende des Kapitels nach, ob Ihre Lösungen richtig sind.

Sie haben in Teilaufgabe d) die Häufigkeiten der einzelnen Klassen ermittelt. Um diese Häufigkeiten graphisch darzustellen, kann man ein *Säulendiagramm* zeichnen. Dazu werden auf der $x$-Achse die Klassen aufgetragen. Über jeder Klasse werden in Form einer Säule die Häufigkeiten aufgetragen. Dabei gibt die Säulenlänge die Häufigkeit der Ausprägung an, d. h. die Anzahl der Elemente, welche in die betreffende Klasse fallen. In

Kapitel 2, »Häufigkeitsverteilungen«, erfahren Sie mehr zur Verwendung von Säulendiagrammen.

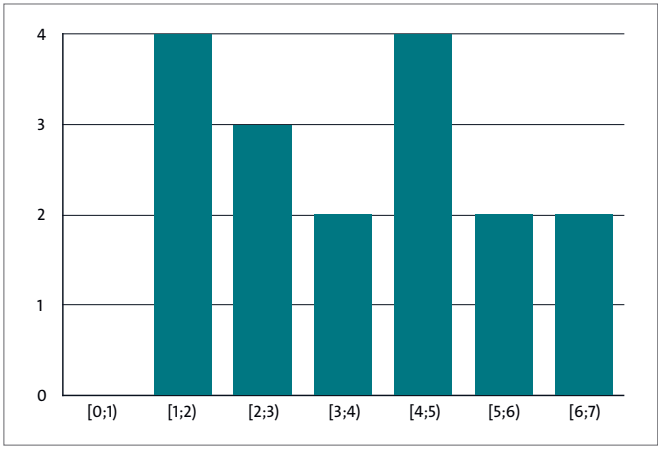



**Abbildung 1.2** Säulendiagramm

### Beispiel 1: Frauenanteil an einem Seminar

Zu einem Seminar melden sich 7 Frauen und 3 Männer an. Kann der hohe Frauenanteil durch Zufall erklärt werden, oder gibt es dafür andere Ursachen wie z. B. eine größere Beliebtheit des Seminarthemas bei Frauen?

Diese Fragestellung muss man der induktiven Statistik zurechnen. Gleichwohl kann man das Beispiel auch ohne induktive Statistik angehen. Dazu simuliert man 1000 Mal je 10 Anmeldungen zu einem Seminar. Jede Anmeldung soll mit gleicher Wahrscheinlichkeit ein männlicher oder ein weiblicher Teilnehmer sein können. Aus der Simulation kann man den Prozentsatz der Simulationen mit 7 oder mehr weiblichen Anmeldungen bestimmen.

Die folgende Abbildung zeigt eine solche Simulation. Männlich und weiblich ist durch *m* und *w* abgekürzt. In jeder Zeile werden 10 Anmeldungen simuliert. Abgebildet sind die ersten 5 der 1000 Simulationen. In dieser speziellen Simulation ergaben sich 181 Fälle mit 7 oder mehr weiblichen Teilnehmern, das sind 18,1 %.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | Anzahl w | Anzahl w>6 | in % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| w | w | m | m | m | m | w | m | w | w | | 5 | **181** | **18,1** |
| w | m | w | m | w | w | w | m | w | w | | 7 | | |
| m | w | w | w | w | w | m | w | m | m | | 6 | | |
| w | w | w | m | w | m | w | w | w | m | | 7 | | |
| w | m | w | m | m | m | w | m | w | m | | 4 | | |

**Abbildung 1.3** Simulation

Weitere Simulationen liefern ähnliche Werte. Wenn ein Ereignis in ca. 18,1 % aller Fälle eintritt, dann kann man es nicht als ungewöhnlich bezeichnen, denn das Ereignis tritt dann etwa in jeder 5. Simulation auf. Das im letzten Beispiel beobachtete Ergebnis ist daher auch durch Zufallsgeschehen erklärbar. Mit der Datei »*1.1 Simulation*« können Sie solche Simulationen selbst durchführen.

**Übungsdateien zu diesem Buch**

Für viele der Beispiele in diesem Buch stehen Excel-Dateien für Sie zur Verfügung, mit denen Sie Simulationen durchführen und mit passendem Datenmaterial praktisch arbeiten können.

Um sie herunterzuladen, scrollen Sie auf der Webseite zum Buch unter *https://www.rheinwerk-verlag.de/5388* etwas herunter bis zu den Materialien zum Buch und verwenden Sie den Link Zu den Materialien.

Hätte die zuvor betrachtete Simulation zu einem Ergebnis geführt, das sich nicht durch ein Zufallsgeschehen beschreiben lässt, dann wäre natürlich noch nicht klar, aus welchen Gründen der Frauenanteil erhöht ist.

**Was Sie wissen sollten**

Sie sollten die Aufgabenfelder der beschreibenden und die der schließenden Statistik kennen.

## 1.2 Wichtige Begriffe

Jede Wissenschaft, die hinreichend weit gekommen ist, benötigt ihre eigene Fachsprache – nicht als Selbstzweck, sondern um die untersuchten Fragen und Probleme exakt beschreiben und zuordnen zu können. Zusätzlich spart eine präzise Fachsprache auch Zeit, weil Sachverhalte ohne Informationsverlust knapper dargestellt werden können. Für die Statistik sind unter anderem die Begriffe *Grundgesamtheit*, *Merkmal*, *Merkmalsträger* und *Merkmalsausprägung* grundlegend. Am Beispiel des Linda-Problems werden diese Begriffe eingeführt. Für die Merkmale wird eine Klassifizierung eingeführt.

### 1.2.1 Das Linda-Problem

»Linda ist 31 Jahre alt, unverheiratet, freimütig und intelligent. Sie hat während ihres Studiums Seminare in Philosophie belegt, interessierte sich als Studentin sehr für Diskriminierung und soziale Ungerechtigkeit und nahm an Demonstrationen gegen Atomwaffen teil.«

Die Kognitionspsychologen Daniel Kahneman und Amos Tversky legten in den 1980er Jahren diese Beschreibung von Linda vielen Versuchspersonen vor. Die Probanden wurden dann gefragt, welche der folgenden Aussagen nach dieser Beschreibung wahrscheinlicher ist:

a) Linda ist bei einer Bank angestellt.

b) Linda ist bei einer Bank angestellt und aktive Feministin.

Die meisten Menschen entscheiden sich für Alternative b) und liegen damit falsch. Kahneman und Tversky legten diese Frage auch Doktoranden des Studiengangs Entscheidungswissenschaft der Stanford Graduate School of Business vor, von denen alle Lehrveranstaltungen zur Wahrscheinlichkeitsrechnung und Statistik besucht hatten. 85 % dieser Probanden hielten »feministische Bankangestellte« für wahrscheinlicher als »Bankangestellte« [KAH].

Der Grund für diese Fehleinschätzung ist, dass unser intuitives Denken ein Faible für plausible Geschichten hat. Je überzeugender eine Geschichte geschildert wird, desto größer ist die Gefahr des hier auftretenden Denkfehlers. Die übliche Bankangestellte ist keine Feministin. Ergänzt man aber das Detail »feministisch« zu »Bankangestellte« – und genau das macht unser Gehirn, ohne uns weiter danach zu fragen –, dann erhält man eine zu den Eingangsinformationen über Linda passende Geschichte.

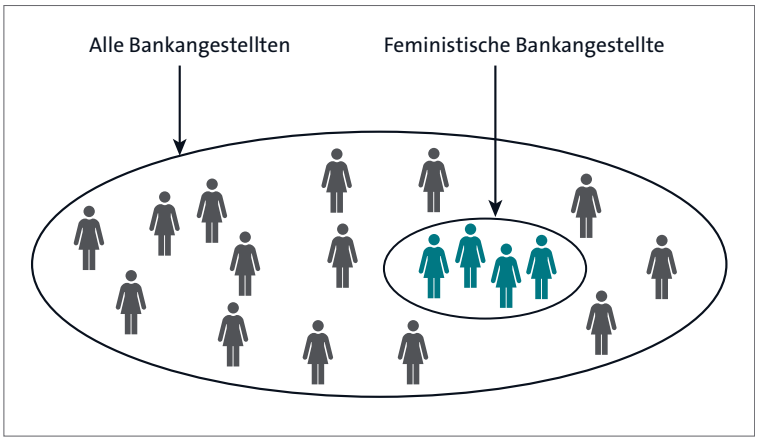



**Abbildung 1.4** Grundgesamtheit im Linda-Problem

Warum aber ist Aussage b) weniger wahrscheinlich? Stellt man sich die Menge aller weiblichen Bankangestellten als Mengendiagramm vor, dann enthält diese Menge die feministischen Bankangestellten als Teilmenge.

Jetzt ist es einsichtig, dass man bei einer zufälligen Auswahl eher eine »Bankangestellte« als eine »feministische Bankangestellte« erhält. Es liegt hier eine immer wieder vorkommende Fehleinschätzung der sogenannten *Grundgesamtheit* vor. Die Grundgesamtheit (auch *Population* oder *statistische Masse*) ist die Menge der »Untersuchungsobjekte«, von denen eine statistische Analyse gemacht werden soll.

### 1.2.2 Merkmale und Merkmalsausprägungen

Als *Merkmale* bezeichnet man in der Statistik die Eigenschaften, die von Interesse sind und die beobachtet werden sollen. Betrachtet man z. B. als *Grundgesamtheit* alle Schüler an deutschen Gymnasien im Schuljahr 2013/2014 – das sind 2 329 990 Schülerinnen und Schüler [STA15] –, dann ist die »Staatsangehörigkeit« ein Merkmal. Das »Geschlecht« ist in diesem Beispiel ein weiteres Merkmal.

Ein Merkmal kann verschiedene *Werte* annehmen. So sind »deutsch« und »italienisch« Werte, welche das Merkmal »Staatsangehörigkeit« annehmen kann. Das Merkmal »Geschlecht« kann üblicherweise die Werte »männlich« und »weiblich« annehmen. Solche Realisierungen eine Merkmals bezeichnet man in der Statistik als *Merkmalsausprägungen*. Ein paar Beispiele sehen Sie in Tabelle 1.1.

| Grundgesamtheit | Merkmal | Merkmalsausprägungen |
|---|---|---|
| Deutsche Städte | Bevölkerung | Insgesamt, männlich, weiblich |
| Seen in Deutschland mit einer Spiegelfläche über 6 $km^2$ | Tiefste Stelle in m | Natürliche Zahlen von 1 bis 300 |
| Schwerbehinderte Menschen | Grad der Behinderung | 50, 60, 70, 80, 90, 100 |
| Männliche Personen ab 65 Jahren | Internetaktivitäten | E-Mails, Soziale Netzwerke, Online Banking, Suche, ... |

**Tabelle 1.1** Beispiele zu Merkmalen und ihren Ausprägungen

Um Aussagen über das Vorkommen bestimmter Merkmalsausprägungen in der Grundgesamtheit zu machen, ist es meist nicht möglich, eine *Totalerhebung*, bei der man die komplette Grundgesamtheit untersucht, durchzuführen. Vielmehr wählt man eine

Teilmenge der Grundgesamtheit aus, die *Stichprobe*. Eine solche Stichprobe sollte *repräsentativ* sein, d. h. sie sollte ein Abbild der Grundgesamtheit sein. Strukturmerkmale wie Einkommen, Geschlecht, Religion und Alter sollten in der Stichprobe und in der Grundgesamtheit gleich vertreten sein. Um das zu erreichen, wählt man die Elemente der Stichprobe zufällig aus der Grundgesamtheit aus. Dadurch hat jedes Element der Grundgesamtheit die gleiche Chance, in die Stichprobe zu gelangen. Man spricht dann von einer *Zufallsstichprobe*.

Manchmal ist eine Zufallsstichprobe aber auch ungeeignet. Will man beispielsweise die Studiendauer von Studentinnen und Studenten des Faches »Deutsch Lehramt« vergleichen und dabei gleich viele Männer wie Frauen befragen, eignet sich eine einer Zufallsstichprobe nicht: Der Frauenanteil in der Stichprobe wäre zu hoch, denn in diesem Studiengang studieren mehr Frauen als Männer. In solchen Situationen setzt man *Quotenstichproben* ein. Dabei werden die Elemente der Stichprobe bewusst ausgewählt, um im angegebenen Beispiel gleich viel Männer und Frauen in der Stichprobe zu haben. Es werden also so lange Elemente aus der Grundgesamtheit gezogen, bis die gewünschte Quote erreicht ist.

Die Quotenstichprobe ist nicht zu verwechseln mit der *geschichteten Zufallsstichprobe*. Bei dieser wird die Grundgesamtheit in einzelne Schichten eingeteilt, und aus jeder Schicht wird dann eine Zufallsstichprobe gezogen. Alle Zufallsstichproben werden dann zusammengeführt. Soll z. B. ermittelt werden, wie zufrieden die Bundesbürger mit ihrer Landesregierung sind, dann ist es naheliegend, die einzelnen Bundesländer als Schichten zu wählen, aus jedem Bundesland eine Zufallsstichprobe zu wählen und diese Stichproben dann zusammenzulegen. Damit ist gewährleistet, dass auch Bürger aus kleineren Bundesländern in der Stichprobe vorkommen.

### 1.2.3 Klassifikation von Merkmalen

Merkmale lassen sich aufgrund der Art der Merkmalsausprägung klassifizieren. So besitzt das Merkmal »Konfession« Merkmalsausprägungen wie katholisch, evangelisch, jüdisch, muslimisch usw., die nicht messbar, d. h. nicht zahlenmäßig vergleichbar sind. Es lässt sich auch keine Rangordnung innerhalb der Merkmalsausprägungen herstellen. Ein solches Merkmal wird *qualitativ* genannt. Bei qualitativen Merkmalen sind die Merkmalsausprägungen durch verbale Ausdrücke gegeben. Das Merkmal »Note« in einer Klassenarbeit hat die Ausprägungen sehr gut, gut, befriedigend, ausreichend, mangelhaft und ungenügend. Hier ist eine Rangfolge vorgegeben. Den Abständen zwischen den Notenstufen entsprechen jedoch nicht gleiche Abstände zwischen den Leistungen. Man kann nicht sagen, dass die Note drei doppelt so gut wie die Note sechs ist.

Auch werden unterschiedliche Leistungen oft in derselben Notenstufe zusammengefasst. Die Noten repräsentieren daher nur die Rangfolge der zugeordneten Schüler. Wegen dieser Rangfolge sind Noten deshalb mehr als ein qualitatives Merkmal. Sie sind ein typisches Beispiel für ein *Rangmerkmal*.

***Qualitatives Merkmal***

Die Merkmalsausprägungen werden **verbal** und nicht durch Zahlen beschrieben.

***Rangmerkmal***

Die Merkmalsausprägungen können in eine Reihenfolge gebracht werden.

***Quantitatives Merkmal***

Die Merkmalsausprägungen sind Zahlen. Falls sich die Menge der Merkmalsausprägungen abzählen lässt, so liegt ein diskretes Merkmal vor, ansonsten ein stetiges Merkmal.

*Quantitative* Merkmale besitzen als Merkmalsausprägung reelle Zahlen. Wenn nur isolierte Zahlenwerte auftreten können, wie z. B. bei der Anzahl der Personen in einem Haushalt oder bei der Kinderzahl in einer Familie, dann spricht man von einem *quantitativ-diskreten* Merkmal. Für *quantitativ-stetige* Merkmale nehmen die Merkmalsausprägungen Werte aus einem Intervall reeller Zahlen an. Beispiele sind Körpergröße, Gewicht oder die zeitliche Dauer eines Telefongesprächs. Genauer lassen sich diskrete und stetige Merkmale mit dem Begriff der *Abzählbarkeit* beschreiben.

***Abzählbarkeit***

Man nennt eine Menge abzählbar, wenn sie aus endlich vielen Elementen besteht oder wenn sie sich mit den natürlichen Zahlen durchnummerieren lässt. Im letzten Fall sagt man dann, dass die Menge abzählbar unendlich ist.

Beispiele für abzählbar unendliche Mengen sind die Menge aller Bruchzahlen oder die Menge der ganzen Zahlen. *Nicht* abzählbar sind die Menge der reellen Zahlen oder auch jedes Intervall $[a, b]$ reeller Zahlen.

***Diskretes Merkmal***

**Diskrete Merkmale** sind dadurch gekennzeichnet, dass die Menge ihrer Merkmalsausprägungen **abzählbar** ist.

***Stetiges Merkmal***

Ein stetiges Merkmal liegt dann vor, wenn die Menge der Merkmalsausprägungen nicht abzählbar ist.

Oft wird bei der Klassifikation der Merkmale nach Art der Skalen unterschieden, die durch die Merkmalausprägungen gebildet werden. Bei qualitativen Merkmalen spricht man von einer *Nominalskala*. Rangmerkmale werden durch *Rangskalen* (oder *Ordinalskalen*) beschrieben. Quantitative Merkmale erzeugen mit ihren Ausprägungen *Kardinalskalen*. Dabei unterscheidet man *Intervallskalen* und *Verhältnisskalen*. Dazu im Folgenden ein paar Beispiele.

»Geschlecht«, »Warenart«, »Blutgruppe (A, B, AB, 0)« und »Studienfach« sind qualitative Merkmale und werden durch eine *Nominalskala* beschrieben. Es gibt keine Reihenfolge der Merkmalsausprägungen. Man kann nur feststellen, ob einzelne Werte gleich oder verschieden sind.

»Platzierung in einem Sportwettbewerb«, »Güteklasse bei Eiern«, »Kleidergröße (S, M, L, XL)« sind Rangmerkmale. Für sie existieren Rangordnungen. Auf einer Ordinalskala können die Merkmalsausprägungen in eine Reihenfolge gebracht werden.

Bei einer Intervallskala liegt nicht nur eine Rangordnung vor, sondern es lässt sich auch der Abstand zwischen Merkmalsausprägungen messen. Beispiele sind »Jahreszahl«, »Datumsangabe« und »Temperatur in Grad Celsius«. Diese Merkmale haben keinen natürlichen Nullpunkt. Der Nullpunkt auf diesen Skalen wurde willkürlich festgelegt (z. B. Schmelzpunkt des Wassers). Deshalb lassen sich zwar Abstände messen, aber keine Verhältnisse bilden. Eine Aussage wie »Heute ist es 5 Grad Celsius wärmer als gestern« ist sinnvoll. Dagegen kann man nicht sagen, dass es bei 10 Grad Celsius doppelt so warm wie bei 5 Grad Celsius wäre. Das sieht man spätestens dann ein, wenn man die Temperaturen in Grad Fahrenheit umwandelt. 5 Grad Celsius entsprechen 41 Grad Fahrenheit, und 10 Grad Celsius entsprechen 50 Grad Fahrenheit. In dieser Skala erscheint die höhere Temperatur nicht als das Doppelte der niedrigen Temperatur. Näheres zur Umrechnung von Grad Celsius in Grad Fahrenheit erfahren Sie in Abschnitt 6.1 ,»Transformationen von Daten«. Einen natürlichen Nullpunkt besitzen z. B. »Kontostand«, »absolute Temperatur in Kelvin« und »Einkommen«. Diese Merkmale werden mit *Ver-*

*hältnisskalen* beschrieben. Hier kann es sinnvoll sein, beispielsweise von einer Verdopplung des Einkommens zu sprechen. In der folgenden Tabelle sind die verschiedenen Skalen und ihre Eigenschaften zusammengestellt. Außerdem sind Beispiele angegeben und die Rechenoperationen, die man mit den Merkmalsauspägungen durchführen kann.

| | **Qualitative Merkmale – Nominalskala** | **Rang-merkmale – Ordinalskala** | **Quantitative Merkmale – Intervallskala** | **Quantitative Merkmale – Verhältnis-skala** |
|---|---|---|---|---|
| **Eigenschaften** | Merkmals-ausprägungen lassen sich nicht ordnen | Merkmals-ausprägungen lassen sich ordnen | Kein natürlicher Null-punkt vorhanden | Natürlicher Nullpunkt vorhanden |
| **Beispiel** | Farbe | Noten | Intelligenz-quotient | Alter |
| **Mögliche Operationen** | $=, \neq$ | $=, \neq, <, >$ | $=, \neq, <, >,$ $+, -$ | $=, \neq, <, >,$ $+, -, \cdot, /$ |

**Tabelle 1.2** Skalen und ihre Eigenschaften

**Aufgabe 2: Merkmale klassifizieren**

Klassifizieren Sie die folgenden Merkmale. Die Klassifikationsgruppen sind: qualitatives Merkmal, Rangmerkmal, quantitativ-diskretes Merkmal und quantitativ-stetiges Merkmal.

Familienstand, gerauchte Zigaretten pro Tag, Güteklasse von Obstsorten, Sehstärke in Dioptrien, Tageshöchsttemperatur, Geburtsland.

### 1.2.4 Zusammenfassung

Im Rahmen eine Datenerhebung sind folgende Begriffe wesentlich:

- *Grundgesamtheit*: Die Menge aller Merkmalsträger
- *Stichprobe*: Eine Teilmenge der Grundgesamtheit, die untersucht wird

- *Merkmalsträger*: Die Objekte der Grundgesamtheit
- *Merkmal*: Die interessierenden Eigenschaften der Merkmalsträger
- *Merkmalsausprägung*: Die Realisierungen eines Merkmals

Die Einteilung von Merkmalen erfolgt in *qualitative Merkmale*, *Rangmerkmale* und *quantitative* Merkmale. Bei den quantitativen Merkmalen werden *diskrete* und *stetige* Merkmale unterschieden.

**Was Sie wissen sollten**

Sie sollten die in diesem Kapitel aufgezählten Grundbegriffe der beschreibenden Statistik kennen.

## 1.3 Lösungen zu den Aufgaben

**Aufgabe 1: Drogeriemarkt-Umsätze**

a) Um den Gesamtumsatz zu ermitteln, addiert man die vorliegenden Zahlen. Es ergibt sich 61,06.

b) Das arithmetische Mittel erhält man, indem man die Summe aller Daten durch die Anzahl der Daten teilt: $61{,}06 / 17 \approx 3{,}59$

c) Dazu ist es sinnvoll, die Liste der Umsätze der Größe nach zu sortieren:

d) 1,02; 1,095; 1,365; 1,47; 2,01; 2,64; 2,7; 3,1; 3,6; 4,2; 4,425; 4,725; 4,8; 5,25; 5,67; 6,15; 6,84

Aus der sortierten Liste kann man die drei kleinsten (1,02; 1,095; 1,365) und die drei größten (5,67; 6,15; 6,84) Umsätze ermitteln.

| Klassen | Häufigkeiten |
|---|---|
| [0; 1) | 0 |
| [1; 2) | 4 |
| [2; 3) | 3 |
| [3; 4) | 2 |
| [4; 5) | 4 |

**Tabelle 1.3** Klassen und ihre Häufigkeiten

| Klassen | Häufigkeiten |
|---|---|
| [5; 6) | 2 |
| [6; 7) | 2 |

**Tabelle 1.3** Klassen und ihre Häufigkeiten (Forts.)

### Aufgabe 2: Merkmale klassifizieren

| Merkmal | Klassifikation |
|---|---|
| Familienstand | qualitativ |
| gerauchte Zigaretten pro Tag | quantitativ-diskret |
| Güteklasse von Obstsorten | Rangmerkmal |
| Sehstärke in Dioptrien | quantitativ-stetig |
| Tageshöchsttemperatur | quantitativ-stetig |
| Geburtsland | qualitativ |

**Tabelle 1.4** Klassifikation von Merkmalen

Kapitel 2

# Häufigkeits-verteilungen

*Die Erhebung von Daten und deren graphische Darstellung sind für die Statistik grundlegend. In diesem Kapitel lernen Sie verschiedene Möglichkeiten kennen, qualitative und ordinale Daten aufzubereiten und darzustellen.*

Zum Einstieg

## 2.1 Darstellung qualitativer und ordinaler Daten

Gummibärchen gibt es in verschiedenen Packungsgrößen zu kaufen. Gängige Größen sind 1 kg, 300 g, 200 g, 100 g und Minibeutel von ca. 10 g. Die Gummibärchen kommen in den Packungen in den sechs verschiedenen Farben Weiß, Gelb, Orange, Hellrot, Dunkelrot und Grün vor, und die Farbzusammenstellungen in den einzelnen Beuteln sind unterschiedlich. Die Herstellerfirma erklärt dies auf folgende Weise:

»Die verschiedenen Mischungsbestandteile eines Artikels werden in einem festen Verhältnis in einen Mischungsbehälter vor der Verpackungsmaschine gefüllt. In diesem Behälter werden die Produktstücke dann vermischt und anschließend als Zufallsmischungen nach Gewicht verpackt. Durch diese Art der Zufallsmischung sind Schwankungen im Mischungsverhältnis unvermeidbar.« [HAR].


**Abbildung 2.1** Gummibärchen

Die Farbzusammenstellung in den Gummibärchenbeuteln scheint von großem Interesse zu sein. Es gibt sogar eine »Gummibärchen Farben Datenbank« im Internet [BAS]. Dort dokumentieren engagierte Gummibärchenkonsumenten die Häufigkeit der Farben in den von ihnen geöffneten Beuteln. Es bleibt anzumerken, dass diese Seite natürlich keinen Anspruch auf wissenschaftliche Ernsthaftigkeit erhebt.

Um die Farbzusammensetzung eines 200-g-Beutels zu erkunden, wird zunächst die sogenannte *Urliste* der Farben erstellt, d. h. für jedes aus dem Beutel entnommene Gummibärchen wird seine Farbe aufgeschrieben.

Urliste: Gelb, Weiß, Grün, Weiß usw.

Mithilfe der Urliste wird ausgezählt, wie oft die einzelnen Merkmalsausprägungen vorkommen. Ein Beispiel dazu sehen Sie in Tabelle 2.1. Eine solche Auflistung nennt man auch eine *Häufigkeitstabelle*.

| Weiß | Gelb | Orange | Hellrot | Dunkelrot | Grün |
|---|---|---|---|---|---|
| 14 | 14 | 15 | 11 | 13 | 16 |

**Tabelle 2.1** Beispiel für Anzahlen der Merkmalsausprägungen, 200-g-Packung

Diese Häufigkeitstabelle enthält die Verteilung der Merkmalsausprägungen in tabellarischer Form. Für die graphische Darstellung der Inhalte kann man ein *Stabdiagramm* verwenden.

Auf der *x*-Achse des Diagramms werden die beobachteten Merkmalsausprägungen aufgetragen. Auf der *y*-Achse werden die Häufigkeiten der Ausprägungen in Form von Stäben eingezeichnet. Die Länge eines jeden Stabes gibt die Häufigkeit der zugehörigen Merkmalsausprägung wieder.

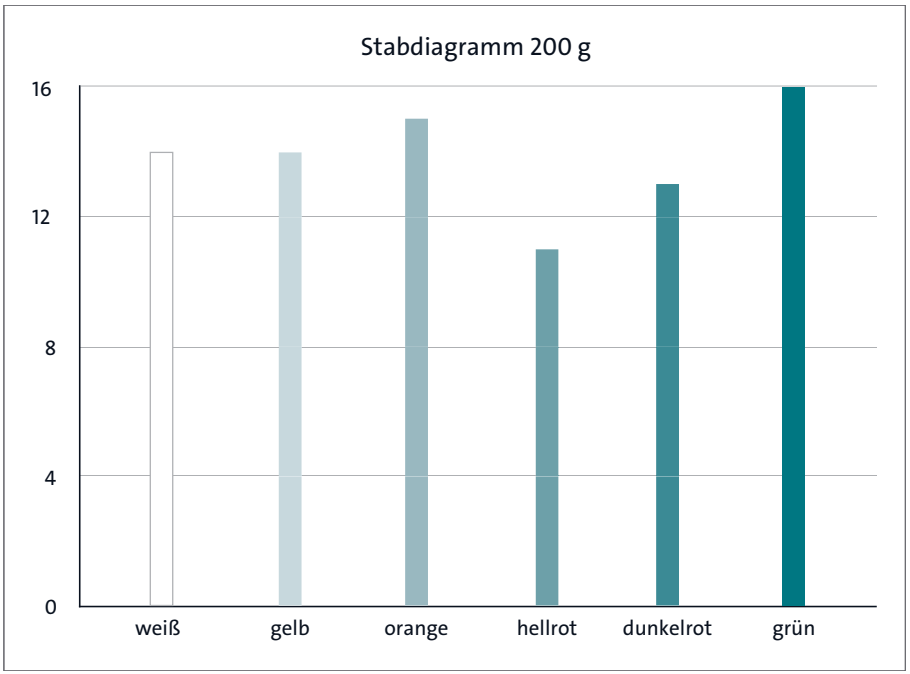



**Abbildung 2.2** Stabdiagramm 200 g

**Aufgabe 1: Farbverteilung von Gummibärchen darstellen**

In einer 300-g-Packung Gummibärchen wurde die folgende Farbverteilung festgestellt:

| Weiß | Gelb | Orange | Hellrot | Dunkelrot | Grün |
|---|---|---|---|---|---|
| 20 | 22 | 27 | 19 | 21 | 22 |

**Tabelle 2.2** Farbverteilung in einer 300-g-Packung

Zeichnen Sie ein Stabdiagramm für diese Verteilung. Sie können dazu auch das Tabellenblatt »*2.1 Stabdiagramm*« in der Datei *2.xlsx* verwenden, mit dem man beliebige Stabdiagramme erstellen kann.

Statt eines Stabdiagramms wird häufig auch ein *Säulendiagramm* benutzt. Dabei ist zu beachten, dass die Breite der Säulen für die Repräsentation der hier vorliegenden quantitativen Daten keine Rolle spielt.

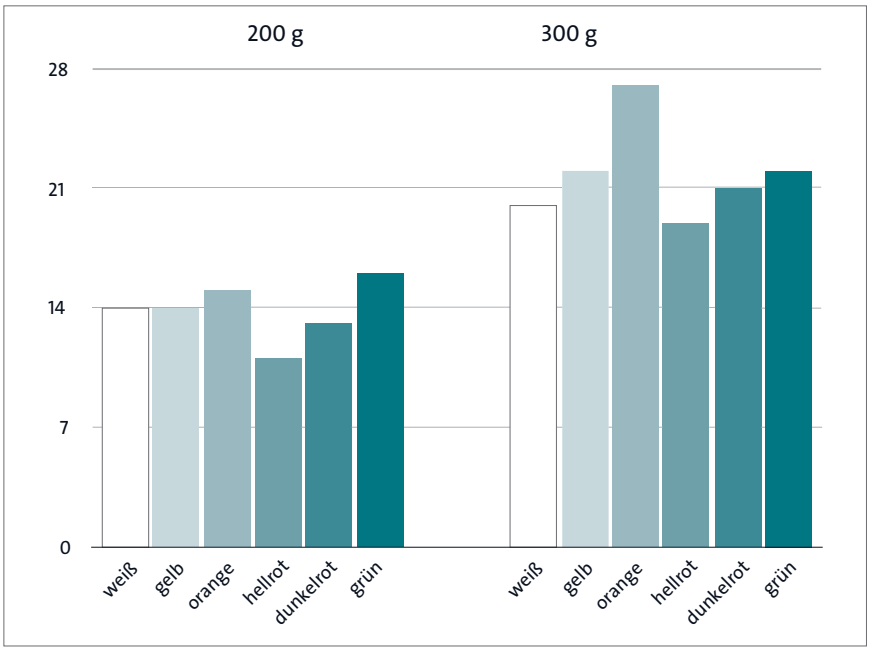



**Abbildung 2.3** Säulendiagramm für 200 g und für 300 g

Säulendiagramme wie in Abbildung 2.3 können Sie mit dem Tabellenblatt »*2.1 Säulendiagramm abs.H*« zeichnen lassen.

Mit den bisher betrachteten Häufigkeiten, die man auch *absolute Häufigkeiten* nennt, ist es nicht möglich, zu entscheiden, ob sich die Farbverteilung im 200-g-Beutel wesentlich von der Verteilung im 300-g-Beutel unterscheidet. Der Grund dafür liegt in der unterschiedlichen Anzahl der Gummibärchen in den beiden Beuteln. Der kleine Beutel enthält 83 Gummibärchen, der große Beutel bringt es auf 131 Gummibärchen. Ein Vergleich zwischen beiden Packungen wird möglich, wenn man für die einzelnen Farben Anteile bildet. Der Anteil der 14 weißen Gummibärchen im 200-g-Beutel ist $14 / 83 \approx 0{,}169 = 16{,}9\,\%$. Der Anteil der 20 weißen Gummibärchen im 300-g-Beutel ist $20 / 131 \approx 0{,}153 = 15{,}3\,\%$. Die auf diese Weise gebildeten Anteile nennt man *relative Häufigkeiten*.

### *Absolute und relative Häufigkeiten*

Die Grundgesamtheit bestehe aus $n$ Merkmalsträgern. Es gebe die Merkmalsausprägungen $a_1, \ldots, a_k$. Dann heißt die Anzahl der Merkmalsträger, bei denen das Merkmal die Ausprägung $a_i$ annimmt, die **absolute Häufigkeit** von $a_i$. Diese Anzahl wird mit $n_i$ bezeichnet.

Dividiert man die absolute Häufigkeit durch den Stichprobenumfang $n$, dann erhält man die **relative Häufigkeit** von $a_i$. Diese wird mit $h_i$ bezeichnet. Es ist $h_i = \frac{n_i}{n}$.

**Beispiel 1: Sitzverteilung im Bundestag**

Nach der Wahl am 24.9.2017 ergab sich die folgende Sitzverteilung im 19. Deutschen Bundestag:

Union: 246, SPD: 153, AfD: 94, FDP: 80, LINKE: 69, GRÜNE: 67

Die Verteilung dieser Daten kann man mit einer Tabelle und einem Diagramm darstellen:

| Merkmalsausprägung | Absolute Häufigkeit | Relative Häufigkeit | Prozent |
|---|---|---|---|
| Union | 246 | 0,347 | 34,7 |
| SPD | 153 | 0,216 | 21,6 |
| AfD | 94 | 0,133 | 13,3 |
| FDP | 80 | 0,113 | 11,3 |

**Tabelle 2.3** Sitzverteilung im 19. Deutschen Bundestag

| Merkmalsausprägung | Absolute Häufigkeit | Relative Häufigkeit | Prozent |
|---|---|---|---|
| Linke | 69 | 0,097 | 9,7 |
| Grüne | 67 | 0,094 | 9,4 |
| **Summen:** | **709** | **1** | **100** |

**Tabelle 2.3** Sitzverteilung im 19. Deutschen Bundestag (Forts.)

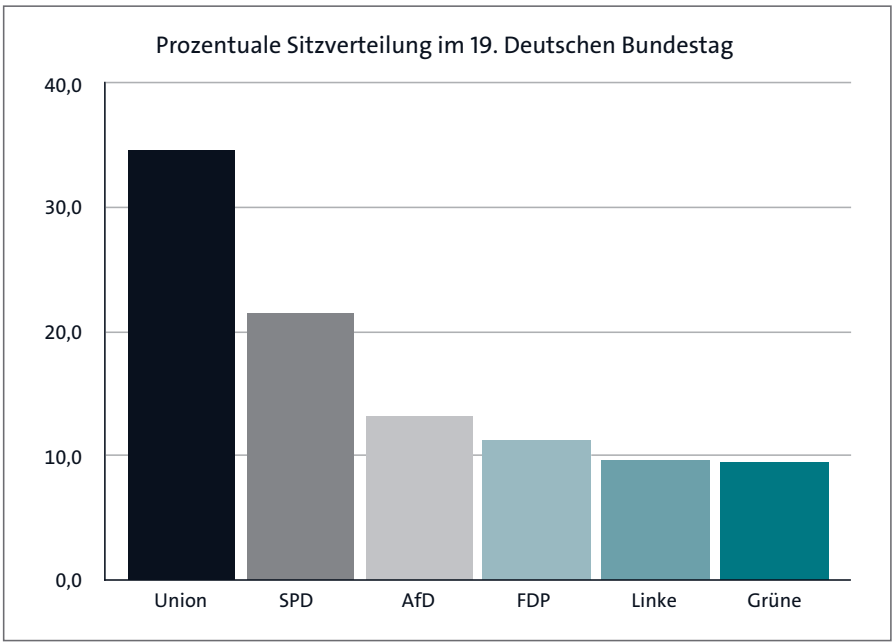



**Abbildung 2.4** Sitzverteilung im Deutschen Bundestag

**Aufgabe 2: Absolute und relative Häufigkeiten untersuchen**

In einer Packung Gummibärchen befinden sich 30 weiße, 36 gelbe, 32 grüne, 30 hellrote, 40 orange und 32 dunkelrote Gummibärchen.

a) Berechnen Sie für jede Farbe die relative Häufigkeit.

b) Bilden Sie die Summe aller absoluten und relativen Häufigkeiten. Welche Eigenschaften der absoluten und der relativen Häufigkeit zeigen sich dabei?

**Merke: Eigenschaften der absoluten und relativen Häufigkeiten**

Die Summe der absoluten Häufigkeiten ist gleich dem Stichprobenumfang $n$. Als Formel geschrieben:

$$\sum_{i=1}^{k} n_i = n$$

Die Summe der relativen Häufigkeiten ist 1. Mit Verwendung des Summenzeichens:

$$\sum_{i=1}^{k} h_i = 1$$

Bei der Formulierung der letzten beiden Eigenschaften wurde das *Summenzeichen* $\Sigma$ verwendet. Dieses wird im weiteren Verlauf noch häufig benötigt werden. Für Leser, denen der Gebrauch des Summenzeichens nicht geläufig ist, gibt es in Abschnitt 2.2 eine Einführung.

**Beispiel 2: Kreisdiagramm zu Haushalten in Deutschland**

Familien, Paare ohne Kinder und Alleinstehende verteilten sich 2014 in Deutschland auf 28 %, 28 % und 44 % der Haushalte [STA15, S. 51].

Dabei umfasst *Familie* alle Eltern-Kind-Gemeinschaften, d. h. Ehepaare, nichteheliche gemischtgeschlechtliche Lebensgemeinschaften, gleichgeschlechtliche Lebensgemeinschaften sowie alleinerziehende Mütter und Väter mit ledigen Kindern im Haushalt. Zur Darstellung dieser relativen Häufigkeiten eignet sich ein *Kreisdiagramm*. In einem Kreisdiagramm wird eine relative Häufigkeit $h_i$ durch den Flächeninhalt einen Kreissektors dargestellt. Dabei ist der Mittelpunktwinkel $\alpha_i$ des Sektors durch $\alpha_i = 360° \cdot h_i$ gegeben.

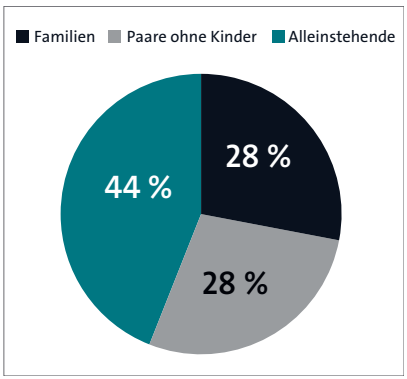



**Abbildung 2.5** Kreisdiagramm

Kreisdiagramme eignen sich nur dann zur Darstellung statistischer Daten, wenn nicht zu viele Kategorien dargestellt werden, denn sonst wird die Darstellung unübersichtlich. Kreisdiagramme können Sie mit Tabellenblatt »*2.1 Kreisdiagramm*« zeichnen lassen.

## 2.2 Das Summenzeichen

Das Summenzeichen Σ (griechisch: Sigma) ist eine bequeme Abkürzung für Summen. Man erspart sich damit die aufwändige Notation vieler Summanden und gewinnt beim Bearbeiten von komplizierten Formeln viel Übersicht.

**Beispiel 3: Arithmetisches Mittel von Körpergrößen**

Von 5 Personen wurde die Körpergröße in cm ermittelt:

| **Person Nr. *i*** | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Größe $x_i$** | 185 | 178 | 191 | 182 | 186 |

**Tabelle 2.4** Körpergrößen

Um auszudrücken, dass die 3. Person 191 cm groß ist, kann man schreiben: $x_3 = 191$

Um die mittlere Körpergröße der 5 Personen zu bestimmen, bildet man das arithmetische Mittel: $\frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5)$

Den Ausdruck in der Klammer kann man jetzt bequem mit dem Summenzeichen schreiben: $\sum_{i=1}^{5} x_i$. Wenn man die Körpergrößen von 100 Personen aufsummieren sollte, so würde man entsprechend $\sum_{i=1}^{100} x_i$ schreiben.

Das Summenzeichen Σ gibt an, dass man die Summe aller rechts von diesem Zeichen stehenden Ausdrücke bildet, wobei man für den sogenannten »Summationsindex«, der unter dem Summenzeichen steht, alle ganzen Zahlen zwischen der unteren und der oberen Grenze einsetzt. Die Zahlen, die man einsetzt, beginnen mit dem Wert unter dem Summenzeichen, der *unteren Grenze* des Summationsindex, und enden mit dem Wert über dem Zeichen, der *oberen Grenze* des Summationsindex.

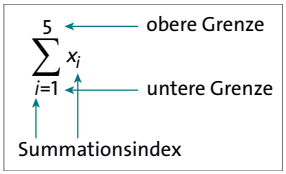



**Abbildung 2.6** Summenzeichen

Statt $x_i$ kann hinter dem Summenzeichen natürlich auch ein anderer Ausdruck stehen.

### Beispiel 4: Summe von Potenzen von 5

In diesem Beispiel steht $5^i$ hinter dem Summenzeichen, d. h. es werden Potenzen von 5 aufsummiert:

$$\sum_{i=1}^{4} 5^i = 5^1 + 5^2 + 5^3 + 5^4 = 5 + 25 + 125 + 625 = 780$$

### Beispiel 5: Summe gerader Zahlen

Die Summe 6 + 8 + 10 + 12 soll mithilfe des Summenzeichens geschrieben werden. Offenbar handelt es sich um aufeinanderfolgende gerade Zahlen, sodass der Ausdruck, der aufsummiert wird, von der Form $2i$ ist. Die untere Grenze von $i$ ist 3, denn 2 mal 3 liefert den ersten Summanden, d. h. die Zahl 6. Die obere Grenze ist 6, denn 2 mal 6 ist gleich dem letzten Summanden. Also kann man diese Summe in folgender Form schreiben: $\sum_{i=3}^{6} 2i$

Man darf sich nicht irritieren lassen, wenn man im Ausdruck hinter dem Summenzeichen einmal keinen Summationsindex findet, wie z. B. bei der Summe $\sum_{i=1}^{5} 1$. Dann ist eben dieser Ausdruck 5-mal zu summieren, sodass gilt: $\sum_{i=1}^{5} 1 = 1 + 1 + 1 + 1 + 1 = 5$

Bisher wurde als Summationsindex immer $i$ benutzt. Das ist rein willkürlich. Man kann den Summationsindex auch umbenennen, was an der Bedeutung der Summe nichts ändert: $\sum_{i=1}^{10} x_i = \sum_{k=1}^{10} x_k$

Vom Rechenunterricht in der Schule kennt man einige Regeln im Umgang mit Summen, z. B. das Ausklammern: $2x_1 + 2x_2 + 2x_3 = 2(x_1 + x_2 + x_3)$

Schreibt man dies mit dem Summenzeichen, dann ergibt sich:

$$\sum_{i=1}^{3} 2 \cdot x_i = 2 \sum_{i=1}^{3} x_i$$

Allgemein formuliert heißt dies:

**Summenregel: Faktoren vorziehen (»ausklammern«)**

Man kann einen Faktor, der nicht vom Summationsindex abhängt, vor das Summenzeichen ziehen:

$$\sum_{i=1}^{n} a \cdot x_i = a \sum_{i=1}^{n} x_i.$$

Vom üblichen Rechnen ist man das Umordnen von Summen gewohnt, z. B.

$(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) = (x_1 + x_2 + x_3) + (y_1 + y_2 + y_3)$.

Mit dem Summenzeichen schreibt sich dies auf folgende Weise:

$$\sum_{i=1}^{3} (x_i + y_i) = \sum_{i=1}^{3} x_i + \sum_{i=1}^{3} y_i$$

Diese Regel lautet dann:

**Summenregel: Summanden umordnen**

Falls hinter dem Summenzeichen eine Summe steht, so kann man die Summation umordnen:

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Die beiden letzten Regeln für das Rechnen mit Summenzeichen lassen sich auch kombinieren.

**Beispiel 6: Summanden und Faktoren**

Hier wird zuerst die Summe umgeordnet, dann werden die vom Summationsindex nicht abhängigen Faktoren vor das Summenzeichen gezogen.

$$\sum_{i=1}^{n} (ax_i + by_i) = \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} by_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i$$

Setzt man dabei $a$ gleich 1 und für $b$ gleich –1 ein, dann erhält man:

$\sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$, d. h. die letzte Regel gilt auch für Differenzen.

Es bleibt darauf hinzuweisen, dass es keine entsprechende Regel für Produkte und Quotienten gibt. Im weiteren Verlauf brauchen Sie diese beiden Summen, die aus dem Schulunterricht bekannt sein könnten:

**Summe der ersten *n* natürlichen Zahlen**

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

**Summenformel der geometrischen Reihe**

$$\sum_{k=1}^{n} x^{k-1} = \frac{1-x^n}{1-x}$$

In manchen Fällen lässt man die Summationsgrenzen am Summenzeichen weg und gibt nur an, wie der Summationsindex heißt. Die zuvor betrachtete Regel über die Umordnung einer Summe hat dann folgendes Aussehen:

$$\sum_{i}(x_i + y_i) = \sum_{i} x_i + \sum_{i} y_i$$

**Aufgabe 3: Regeln für Summen anwenden**

Ordnen Sie die folgende Summe in Einzelsummen um, und vereinfachen Sie sie mithilfe der zuvor betrachteten Regeln:

$$\sum_{i=1}^{n}(ax_i + by_i + c)$$

**Was Sie wissen sollten**

Sie sollten Stabdiagramme, Säulendiagramme und Kreisdiagramme als Darstellungsformen kennen.

Sie sollten absolute und relative Häufigkeiten verwenden können.

Sie sollten die Eigenschaften der relativen Häufigkeiten kennen.

Sie sollten das Summenzeichen verwenden können.

## 2.3 Darstellung quantitativ-diskreter Daten

In diesem Abschnitt sehen Sie, wie Sie Häufigkeitstabellen und Diagramme auch für quantitativ-diskrete Daten verwenden.

**Beispiel 7: Wie viele Geschwister?**

In einer Klasse von 30 Schülern wird für jeden Schüler die Anzahl der Geschwister festgestellt. Das ergibt die folgende Urliste:

0, 3, 2, 2, 0, 1, 0, 0, 1, 1, 1, 0, 2, 1, 0, 0, 0, 0, 1, 1, 2, 1, 1, 0, 0, 1, 0, 0, 1, 0

| Geschwisterzahl | Absolute Häufigkeit | Relative Häufigkeit |
|---|---|---|
| 0 | 14 | 0,467 |
| 1 | 11 | 0,367 |
| 2 | 4 | 0,133 |
| 3 | 1 | 0,033 |

**Tabelle 2.5** Relative Häufigkeit der Geschwisterzahl

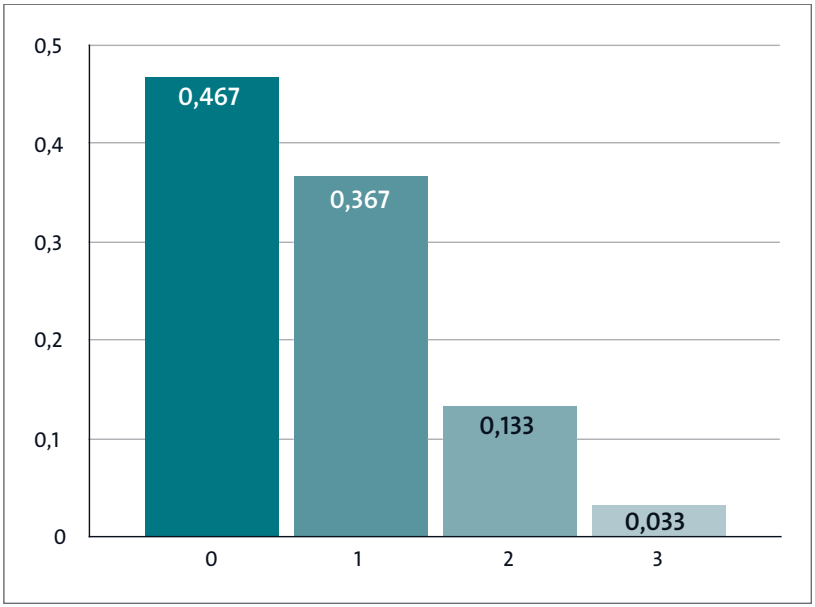



**Abbildung 2.7** Relative Häufigkeit der Geschwisterzahl

In Beispiel 7 treten nur die Merkmalsausprägungen 0, 1, 2 und 3 auf. Die Darstellung unterscheidet sich von einer Häufigkeitstabelle und einem Diagramm für ein qualitatives Merkmal oder ein Rangmerkmal.

Enthält die Urliste der Daten aber viele Merkmalsausprägungen, dann ist es sinnvoll, eine *Klassenbildung* vorzunehmen, d. h. einzelne Merkmale zu einer Klasse zusammenzufassen. Dabei dürfen sich die einzelnen Klassen nicht überschneiden. Die absolute Häufigkeit einer Klasse gibt an, wie oft eine Merkmalsausprägung in dem Datensatz auftritt, der zwischen den Grenzen dieser Klasse liegt. Die relativen Häufigkeiten erhält man, indem man die absoluten Klassenhäufigkeiten durch die Anzahl der Daten teilt.

**Beispiel 8: Migrationshintergrund und Lebensalter**

Die Altersverteilung der Bevölkerung mit Migrationshintergrund im Alter von 15 Jahren (einschließlich) bis 65 Jahren (ausschließlich) soll dargestellt werden.

In diesem Fall ist es nicht sinnvoll, ein Diagramm mit 50 Säulen zu zeichnen. Vielmehr werden die Daten in Klassen der Breite 10 zusammengefasst. Dabei deutet die linke eckige Klammer an, dass diese Grenze zur Klasse gehört, die rechte runde Klammer sagt aus, dass diese Grenze *nicht* zu der Klasse gehört [STA16, S. 80]. Die Tabelle 2.6 enthält zu jeder der angegebenen Klassen deren absolute und relative Häufigkeit.

| Klasse | Absolute Anzahl $n_i$ in der Einheit 1 000 | Relative Häufigkeit $h_i$ |
|---|---|---|
| [15; 25) | 2 262 | 0,1932 |
| [25; 35) | 2 620 | 0,2238 |
| [35; 45) | 2 764 | 0,2361 |
| [45; 55) | 2 327 | 0,1988 |
| [55; 65) | 1 735 | 0,1482 |
| | Summe: $n$ = 11 708 | |

**Tabelle 2.6** Altersverteilung

Abbildung 2.8 zeigt ein Säulendiagramm, in dem jeder Klasse deren relative Häufigkeit zugeordnet ist.

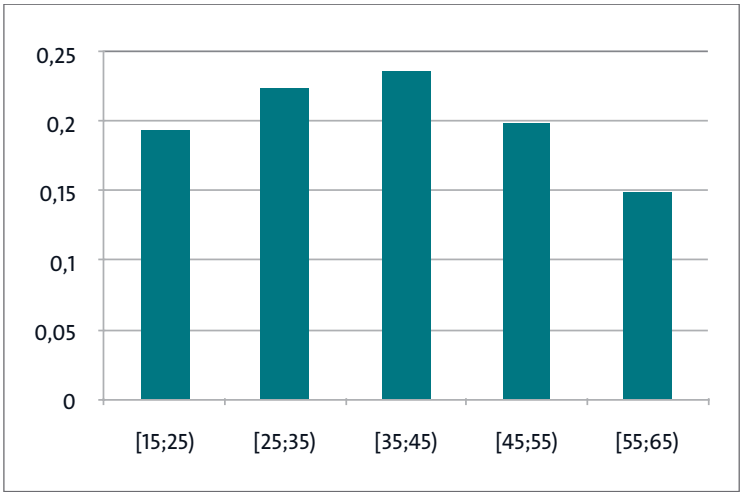



**Abbildung 2.8** Klasseneinteilung d. Bevölkerung mit Migrationshintergrund

Für quantitativ-diskrete Merkmale gibt es weitere Darstellungsmöglichkeiten, die zum Beispiel deutlich machen, welcher Anteil der betrachteten Bevölkerung jünger als 25 oder jünger als 55 Jahre ist. Sie werden sie in Abschnitt 2.5.1 kennen lernen.

**Aufgabe 4: Mietpreise für 1-Zimmer-Wohnungen darstellen**

Auf einer Immobilienseite im Internet gab es im März 2017 folgende Angebote für 1-Zimmer-Wohnungen in Mannheim (alle Angaben in Euro): 310, 480, 260, 270, 350, 260, 480, 500, 680, 430, 320, 295, 370, 300, 300, 360, 250, 410, 450, 300.

Klassieren Sie diese Daten in Klassen mit der Klassenbreite 50, wobei die erste Klasse bei 250 beginnt. Erstellen Sie eine Häufigkeitstabelle für die relativen Klassenhäufigkeiten, und zeichnen Sie ein Säulendiagramm.

**Was Sie wissen sollten**

Sie sollten wissen, dass Sie für quantitativ-diskrete Merkmale alle Darstellungsformen verwenden können, die auch für qualitative Merkmale möglich sind.

Sie sollten für quantitativ-diskrete Merkmale Klasseneinteilungen und deren Häufigkeitstabellen erstellen können.

## 2.4 Darstellung quantitativ-stetiger Daten

Vom 3. bis zum 5.2.2017 fand auf der umgebauten Heini-Klopfer-Skiflugschanze ein Weltcup-Springen statt, das als Generalprobe für die Skiflugweltmeisterschaft im Jahr 2018 galt. Der Deutsche Andreas Wellinger stellte dabei am 5.2.2017 den Schanzenrekord von 238 m auf. Am Vortag fand ein Einzelwettbewerb mit 30 Springern statt. Jeder Springer absolvierte zwei Sprünge. Sieger wurde der Österreicher Stefan Kraft vor Andreas Wellinger. Kraft war zwar etwas kürzer als Wellinger geflogen, aber dafür besser gelandet. Die Sprungweiten aller $n = 60$ Sprünge sind in der Tabelle zusammengefasst, die Namen der Sportler wurden weggelassen.

| 1. Sprung | 2. Sprung | 1. Sprung | 2. Sprung | 1. Sprung | 2. Sprung |
|---|---|---|---|---|---|
| 227,5 | 218 | 208,5 | 210,5 | 181,5 | 196 |
| 234,5 | 222,5 | 193 | 210 | 190 | 204 |
| 222,5 | 217 | 185,5 | 216 | 188 | 201,5 |
| 215,5 | 229 | 195 | 219,5 | 181,5 | 180,5 |
| 220 | 211 | 217,5 | 197 | 171,5 | 177,5 |
| 220 | 211,5 | 188,5 | 220,5 | 186,5 | 186,5 |
| 207 | 215 | 205,5 | 210,5 | 185 | 185,5 |
| 211,5 | 214,5 | 188,5 | 204 | 183,5 | 188 |
| 214 | 203 | 185 | 214,5 | 185,5 | 162,5 |
| 195,5 | 214 | 183 | 215,5 | 169,5 | 179,5 |

**Tabelle 2.7** Sprungweiten

Die Sprungweiten der Springer sind reelle Zahlen und stellen damit ein stetiges Merkmal dar. Wenn man ganz genau messen würde, dann wäre es möglich, zu zeigen, dass es keine zwei Sprünge gibt, welche exakt die gleiche Weite haben. Aus diesem Grund ergibt es bei einem stetigen Merkmal keinen Sinn, für jede Merkmalsausprägung die relative Häufigkeit zu erheben, wie das zuvor für qualitative Daten erfolgt ist. Jede Merkmalsausprägung käme nur einmal vor, und alle Merkmalsausprägungen hätten deshalb die relative Häufigkeit 1/60.

Tatsächlich ist die Messung der Weiten aber nur im Halbmeter-Maßstab erfolgt, sodass eigentlich qualitativ-diskrete Daten vorliegen. Wegen der vielen unterschiedlichen Merkmalsausprägungen werden die Daten trotzdem wie stetige Daten behandelt.

Mit den vorliegenden Daten soll die Verteilung der Sprungweiten dargestellt werden. Das wird dadurch ermöglicht, dass man eine Klasseneinteilung durchführt, im vorliegenden Beispiel mit der Klassenbreite $d = 5$ (Einheit Meter). Es werden 15 Klassen gebildet: [160; 165), [165; 170), ..., [230; 235). Die Klammerung der Klassen soll andeuten, dass die linke Grenze zur jeweiligen Klasse gehört, die rechte Grenze dagegen nicht. Für jede Klasse wird die Klassenhäufigkeit $n_i$, d. h. die Anzahl der Weiten, die in diese Klasse fallen, ermittelt und in die dritte Spalte der Tabelle 2.8 eingetragen. Teilt man die Klassenhäufigkeiten $n_i$ durch die Anzahl $n$ der Sprünge, dann ergibt sich für jede Klasse die relative Klassenhäufigkeit $h(x_i)$. Diese wird in die 4. Spalte der Tabelle 2.8 eingetragen.

| Klassen | Klassenmitte $x_i$ | Häufigkeit $n_i$ | Rel. Häufigkeit $h(x_i) = n_i / n$ | Säulenhöhe $h(x_i) / d$ |
|---|---|---|---|---|
| [160; 165) | 162,5 | 1 | 0,01667 | 0,00333 |
| [165; 170) | 167,5 | 1 | 0,01667 | 0,00333 |
| [170; 175) | 172,5 | 1 | 0,01667 | 0,00333 |
| [175; 180) | 177,5 | 2 | 0,03333 | 0,00667 |
| [180; 185) | 182,5 | 5 | 0,08333 | 0,01667 |
| [185; 190) | 187,5 | 11 | 0,18333 | 0,03667 |
| [190; 195) | 192,5 | 2 | 0,03333 | 0,00667 |
| [195; 200) | 197,5 | 4 | 0,06667 | 0,01333 |
| [200; 205) | 202,5 | 4 | 0,06667 | 0,01333 |
| [205; 210) | 207,5 | 3 | 0,05 | 0,01 |
| [210; 215) | 212,5 | 10 | 0,16667 | 0,03333 |
| [215; 220) | 217,5 | 8 | 0,13333 | 0,02667 |
| [220; 225) | 222,5 | 5 | 0,08333 | 0,01667 |
| [225; 230) | 227,5 | 2 | 0,03333 | 0,00667 |
| [230; 235) | 232,5 | 1 | 0,01667 | 0,00333 |

**Tabelle 2.8** Klasseneinteilung der Sprungweiten

Man möchte nun die relativen Klassenhäufigkeiten durch Rechtecksäulen graphisch darstellen. Dabei stellt die Breite einer Säule die Klassenbreite $d$ dar. Die Höhe der Säule wird so gewählt, dass der Flächeninhalt der jeweiligen Säule als relative Häufigkeit der Klasse interpretiert werden kann. Dazu muss die Säulenhöhe gleich $\frac{h(x_i)}{d}$ sein. Diese Werte sind in der letzten Spalte der Tabelle 2.8 eingetragen.

Bildet man jetzt den Flächeninhalt einer Säule durch die Rechnung »Breite mal Höhe«, das heißt $d \cdot \frac{h(x_i)}{d}$, dann ergibt sich, wie gewünscht, $h(x_i)$. Man bezeichnet die Rechteckhöhe auch als *Häufigkeitsdichte* $f(x_i)$.

Für die Skiflugdaten aus Tabelle 2.8 wurden diese Rechtecksäulen gezeichnet. Die sich ergebende Figur heißt das *Histogramm* der Verteilung.

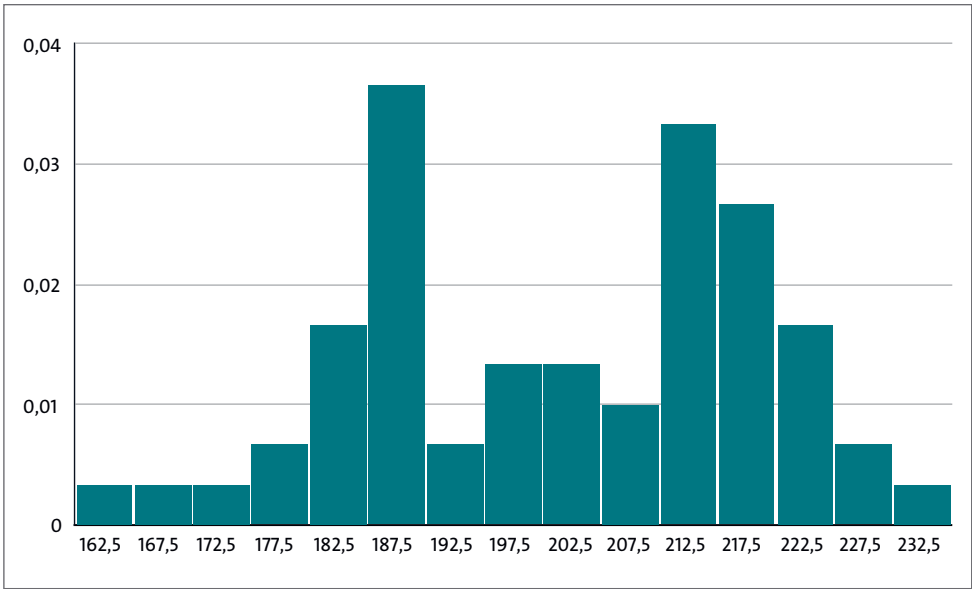



**Abbildung 2.9** Histogramm Skiflug Obersdorf 4.2.2017

Das Histogramm zeigt, dass sich die Flugweiten in zwei Gruppen einteilen lassen, und zwar in Weiten zwischen 160 m und 200 m und in Weiten zwischen 200 m und 235 m. Dem entspricht auch, wie man in der Tabelle mit den ersten und zweiten Sprüngen sehen kann, das Leistungsvermögen der einzelnen Springer. Meist unterscheiden sich der erste und der zweite Sprung eines Springers nicht allzu sehr in der Weite. Die Klassen-

einteilung hat es ermöglicht, festzustellen, dass die Springer zwei Leistungsgruppen zugeordnet werden können.

Hier wird zusammengefasst, was man unter der Häufigkeitsdichte versteht:

### *Häufigkeitsdichte*

Betrachtet wird ein quantitativ-stetiges Merkmal, für das $n$ Merkmalsausprägungen und eine Klasseneinteilung mit der konstanten Klassenbreite $d$ vorliegen. Mit $x_i$ ist die Klassenmitte der $i$-ten Klasse bezeichnet. Die Zahl der Merkmalswerte, welche in die $i$-te Klasse fallen, ist $n_i$. Mit $h(x_i) = n_i / n$ wird die relative Klassenhäufigkeit bezeichnet. Dann nennt man

$$f(x_i) = \frac{n_i}{n \cdot d} = \frac{h(x_i)}{d}$$ die **Häufigkeitsdichte**.

Die Häufigkeitsdichte wird als **Rechteckhöhe** in einem Histogramm dargestellt. Der Flächeninhalt des Rechtecks stellt die relative Klassenhäufigkeit dar.

Mithilfe der Häufigkeitsdichte kann man die Summe aller Rechteckflächen bestimmen. Der Flächeninhalt des $i$-ten Rechtecks ist gleich $f(x_i) \cdot d$, denn $f(x_i)$ ist die Rechteckhöhe und $d$ die Breite. Die Summe aller Rechteckflächen ist dann:

$$\sum_i f(x_i) \cdot d = \sum_i \frac{h(x_i)}{d} \cdot d = \sum_i h(x_i) = 1$$

Dabei wurde die Eigenschaft der relativen Häufigkeiten, dass ihre Summe 1 ergibt, verwendet. Ein Histogramm besitzt also die folgenden Eigenschaften:

**Merke: Eigenschaften eines Histogramms**

Die relative Häufigkeit einer Klasse ist gleich dem Flächeninhalt des zugehörigen Rechtecks des Histogramms.

Die Summe aller Rechteckflächen des Histogramms ist 1.

Wenn Sie selbst ein Histogramm erstellen müssen, dann können Sie sich am folgenden Ablaufplan orientieren.

**Wie Sie ein Histogramm erstellen können**

Ein Histogramm ist die graphische Darstellung der Häufigkeitsverteilung eines quantitativ-stetigen Merkmals durch Rechtecke. Um das Histogramm zu zeichnen, gehen Sie folgendermaßen vor:

1. Sie müssen die vorliegenden $n$ Daten in Klassen der Breite $d$ einteilen (ca. 5 bis 25 Klassen; oft verwendete Faustregel: Wurzel aus $n$ Klassen). Die Klassen dürfen sich dabei nicht überschneiden.
2. Bestimmen Sie zu jeder Klasse die Klassenmitte $x_i$.
3. Zählen Sie für jede Klasse aus, wie viele Elemente $n_i$ in diese Klasse fallen.
4. Ermitteln Sie die relativen Klassenhäufigkeiten $h_i$, d. h. teilen Sie die Zahlen $n_i$ durch die Anzahl $n$ aller Daten. Rechenkontrolle: Die Summe aller $h_i$ muss 1 ergeben. (Wenn Sie mit Dezimalzahlen statt mit Brüchen rechnen, verhindern Rundungsfehler manchmal, dass sich genau 1 ergibt.)
5. Berechnen Sie die Höhe der Rechtecke, indem Sie die Zahlen $h_i$ durch die Klassenbreite $d$ teilen.
6. Zeichnen Sie für jede Klasse ein Rechteck mit der Breite $d$ und der zuletzt berechneten Höhe.

**Aufgabe 5: Ein Histogramm zu Blutdruckwerten erstellen**

Blutdruckwerte werden durch zwei Zahlen angegeben, den systolischen und den diastolischen Blutdruck. In unbelastetem Zustand ist ein Blutdruckwert, der 120/80 mmHg nicht überschreitet, optimal. Die Tabelle 2.9 gibt den systolischen, d. h. den höheren Blutdruckwert einer Person an 50 aufeinander folgenden Tagen an.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 138 | 122 | 144 | 134 | 142 | 131 | 117 | 127 | 143 | 138 |
| 116 | 119 | 133 | 148 | 135 | 141 | 137 | 120 | 128 | 137 |
| 120 | 119 | 145 | 144 | 140 | 128 | 125 | 133 | 136 | 137 |
| 123 | 129 | 127 | 138 | 144 | 126 | 146 | 121 | 136 | 143 |
| 139 | 126 | 122 | 133 | 148 | 132 | 125 | 127 | 133 | 153 |

**Tabelle 2.9** Blutdruckwerte

Die Werte sind zwar als ganze Zahlen angegeben, trotzdem handelt es sich beim Blutdruck um ein stetiges Merkmal. Die Daten schwanken von 116 bis 153. Es liegen 38 verschiedene Merkmalsausprägungen vor. Führen Sie eine Klasseneinteilung mit Klassen der Breite 7 durch. Beginnen Sie bei 115, und enden Sie bei 157. Berechnen Sie für jede der Klassen die Häufigkeitsdichte, und zeichnen Sie das zugehörige Histogramm.