Einführung in die Inferenzstatistik

Dieses Buch wurde mit der Zielsetzung geschrieben, Medizinern oder Psychologen bei der statistischen Auswertung klinisch-empirischer Untersuchungen behilflich zu sein. Wenn hier von statistischer Auswertung die Rede ist, meinen wir nicht die deskriptive Aufarbeitung der erhobenen Daten (Laborwerte, Symptome, Diagnosen, Krankheitsverläufe etc.) in Form von zusammenfassenden Grafiken oder statistischen Kennwerten (Mittelwerte, Streuungen etc.). Dies ist Aufgabe der beschreibenden Statistik, die hier nicht thematisiert wird. Unser Anliegen ist es, ein einfaches Instrumentarium bereitzustellen, mit dem man herausfinden kann, ob eine klinische Forschungshypothese durch eine empirische Untersuchung bestätigt wird oder nicht. Wenn die Anzahl der untersuchten Patienten relativ klein ist, sind hierfür die sog. verteilungsfreien Tests besonders geeignet, die Gegenstand dieses Buches sind (▶ Abschn. 1.2.6).

Angenommen, man hätte die Hypothese formuliert, eine neue Behandlungsmethode A sei erfolgreicher als eine alte bewährte Behandlungsmethode B, und eine empirische Untersuchung möge zeigen, dass in der Tat mit der neuen Methode 10% mehr Behandlungserfolge erzielt werden als mit der alten Methode. Kann man nun davon ausgehen, die Überlegenheit der neuen Methode sei erwiesen oder gar „bewiesen"? Nehmen wir ferner an, man habe mit beiden Methoden jeweils 20 Patienten behandelt mit 10 Behandlungserfolgen nach der alten Methode. „10% mehr" von 10 Behandlungserfolgen bedeuten nichts anderes, als dass mit der neuen Methode 11 Patienten, also lediglich ein Patient mehr erfolgreich behandelt werden konnte. Müssen wir uns angesichts dieser Zahlen nicht fragen, ob diese „Überlegenheit" nichts anderes ist als ein Produkt des Zufalls?

Diese Thematik, die Absicherung eines Untersuchungsergebnisses gegen ein Zufallsergebnis, ist zentral für alle statistischen Verfahren zur Hypothesenprüfung *(Inferenzstatistik)*.

Die Frage, ob ein empirisches Untersuchungsergebnis zufallsbedingt sein kann oder nicht, sollte in jeder klinischen Forschungsarbeit beantwortet werden. Die Vielfalt der Fragestellungen, die Gegenstand einer statistischen Hypothesenprüfung sein können, sei anhand einiger Beispiele, die in den folgenden Kapiteln ausführlich behandelt werden, exemplarisch verdeutlicht:

- Ist die erhöhte Krebsmortalität in einem Wohnhaus mit dem Zufall zu erklären, oder sind hierfür besondere, krebsfördernde Ursachen verantwortlich zu machen? (▶ Beispiel 2.1)
- Ist die Behandlung schizophrener Patienten mit einem typischen Neuroleptikum erfolgreicher als mit einem atypischen Neuroleptikum, oder sind die Wirkunterschiede u. U. nur zufallsbedingt? (▶ Beispiel 3.3)
- Wird die Lebensqualität von Patienten mit Coxarthrose durch eine Endoprothese tatsächlich „überzufällig" verbessert? (▶ Beispiel 4.2)
- Muss man davon ausgehen, dass der Zusammenhang zwischen dem Verlaufsstadium von chronisch-obstruktiven Lungenerkrankungen und dem Schweregrad der krankheitsbedingten Dyspnoe zufallsbedingt ist, wenn man das Alter der Patienten berücksichtigt? (▶ Beispiel 5.6)
- Sind Internisten in der Lage, die Ätiologie einer chronischen Hepatitis übereinstimmend zu diagnostizieren, oder sind die erzielten Übereinstimmungen ein Produkt des Zufalls? (▶ Beispiel 6.2)
- Können die Schwankungen der täglich gemessenen Blutzuckerwerte bei einem Patienten mit Diabetes mellitus Typ II zufallsbedingt sein, oder verbirgt sich hinter den Schwankungen eine Systematik? (▶ Beispiel 8.2)

Jede dieser Fragestellungen erfordert einen eigenständigen Hypothesentest. Die in diesem Buch behandelten Tests sind vorne im Buchdeckel in einer nach Art der Fragestellung und Datenart gegliederten Übersicht zusammengestellt.

Die Hypothesentests machen letztlich nichts anderes, als die Wahrscheinlichkeit zu ermitteln, mit der das Untersuchungsergebnis ein reines Zufallsergebnis ist. Wenn diese Wahrscheinlichkeit sehr klein ist, können wir vermuten, dass das Untersuchungsergebnis *nicht* zufallsbedingt ist, sondern einen systematischen Effekt anzeigt. Hierüber werden wir im ▶ Abschn. 1.2 ausführlicher berichten.

Zuvor jedoch müssen wir uns mit einigen grundlegenden Begriffen auseinandersetzen, die für alle statistischen Hypothesentests zentral sind. Hierzu gehört insbesondere der Wahrscheinlichkeitsbegriff, mit dem wir uns im Folgenden beschäftigen.

1.1 Zum Begriff „Wahrscheinlichkeit“

Wir alle kennen das auf die beschreibende Statistik gemünzte Wort: „Mit Statistik kann man alles beweisen!“ Richtiger müsste es aus dem Blickwinkel der hypothesenprüfenden Statistik heißen: Mit Statistik kann man gar nichts beweisen, keinen Unterschied, keinen Zusammenhang, keine Gesetzmäßigkeit, sofern man von einem Beweis fordert, dass er logisch und sachlich unwidersprochen bleiben soll.

Was kann die moderne Statistik als wissenschaftliche Methode wirklich leisten? Sie gibt Auskunft darüber, mit welcher Wahrscheinlichkeit Unterschiede, Zusammenhänge und Regelmäßigkeiten, die wir in Stichprobenerhebungen gefunden haben, rein zufällig entstanden sein können, oder inwieweit sie als allgemein gültig anzusehen sind. Absolut sichere Aussagen und Voraussagen sind mithilfe der Statistik unmöglich. Jedoch liegt es an uns, das Risiko bzw. die Wahrscheinlichkeit dafür, dass unsere Aussage falsch ist, nach Art der wissenschaftlichen Fragestellung höher oder niedriger anzusetzen.

Der Begriff „Wahrscheinlichkeit“ ist uns auch im Alltag geläufig. Wenn beispielsweise im Wetterbericht bekanntgegeben wird, dass es heute mit einer Wahrscheinlichkeit von 90% regnen wird, dürfte es wohl kaum jemand versäumen, für den geplanten Spaziergang einen Regenschirm mitzunehmen. Auch Fragen nach der Wahrscheinlichkeit, beim Münzwurf „Zahl“ zu werfen, mit einem Würfel eine Sechs zu würfeln oder aus einem Skatspiel mit 32 Karten zufällig das Herz-As zu ziehen, verlangen keine besonderen mathematisch-statistischen Kenntnisse. Für die Beantwortung der letztgenannten Fragen gibt es eine einfache Regel, die generell für gleich wahrscheinliche bzw. gleichwertige Ereignisse gilt: Wir überlegen uns die Anzahl der „günstigen“ Ereignisse (dies sind die Ereignisse, deren Wahrscheinlichkeit wir bestimmen wollen) und dividieren diese Anzahl durch die Anzahl aller möglichen Ereignisse (dies sind alle Ereignisse, die im jeweiligen Versuch vorkommen können). Das Resultat ist die gesuchte Wahrscheinlichkeit, die wir mit dem Buchstaben p symbolisieren (von „probabilité“).

$$p = \frac{\text{Anzahl der günstigen Ereignisse}}{\text{Anzahl der möglichen Ereignisse}} \tag{1.1}$$

Für das Münzbeispiel erhalten wir nach dieser Regel $p=1/2$ (oder 50%), für das Würfelbeispiel $p=1/6$ (oder 16,7%) und für das Skatbeispiel $p=1/32$ (oder 3,1%).

Mit Gl. 1.1 können wir beispielsweise auch bestimmen, wie groß die Wahrscheinlichkeit ist, mit einem Würfel eine ungerade Zahl zu werfen. Da hier die Zahlen 1, 3 und 5 „günstige" Ereignisse darstellen, und 6 Zahlen möglich sind, erhält man $p=3/6=1/2$. In gleicher Weise ermittelt man die Wahrscheinlichkeit für eine Pik-Karte im Skatspiel zu $8/32=1/4$, da 8 der 32 Skatkarten die „Farbe" Pik aufweisen.

Aus Gl. 1.1 geht hervor, dass jede Wahrscheinlichkeit einen Wert p hat, der nicht negativ und nicht größer als 1 ist, d. h. die Wahrscheinlichkeitsskala erstreckt sich von 0 (unmögliches Ereignis) bis 1 (sicheres Ereignis).

Wir haben die obige Definition noch etwas näher zu erläutern. Halten wir uns dabei an das Würfelbeispiel: Die Anzahl der möglichen Ereignisse beträgt 6. Diese 6 Ereignisse schließen einander aus, denn man kann nicht sowohl eine 4 als auch eine 6 im selben Wurf erzielen. Die 6 Ereignisse sind auch gleichwertig, denn jedes Ergebnis hat die gleiche Chance aufzutreten.

Die beiden Begriffe „gleichwertig" und „einander ausschließend" wollen wir an 2 Beispielen illustrieren.

Beispiel 1.1. Kartenspiel

Jemand möchte die Wahrscheinlichkeit, aus einem Skatspiel entweder ein As oder eine Herz-Karte zu ziehen, ermitteln. Das Kartenspiel enthält 32 Karten, darin befinden sich 4 Asse und 8 Herz-Karten. Folglich stehen – so möchte man meinen – die günstigen Ereignisse im Verhältnis zu den möglichen Ereignissen wie 12:32, also ist $p=0{,}375$. Diese Schlussfolgerung ist aber unrichtig, denn ein As (das Herz-As) gilt zugleich auch als Herz-Karte. Das Auftreten eines Asses schließt also das Auftreten einer Herz-Karte nicht aus. Die Bedingung, dass die Ereignisse einander ausschließen sollen, ist nicht erfüllt. Daher sind wir zu einem unrichtigen Wahrscheinlichkeitswert gekommen. Der richtige beträgt $p=11/32=0{,}344$.

Beispiel 1.2. Münzwurfspiel

Angenommen, jemand möchte die Wahrscheinlichkeit ermitteln, bei 2 hintereinander durchgeführten Würfen mit einer Münze 2-mal Zahl zu erhalten. Die 3 möglichen Ergebnisse, 2-mal Zahl, 2-mal Adler sowie einmal Zahl und einmal Adler schließen sich gegenseitig aus. Man könnte also schlussfolgern, die Wahrscheinlichkeit, 2-mal Zahl zu werfen, betrage 1/3. Diese Überlegung ist falsch, denn die 3 Ereignisse sind nicht gleichwertig. Das 3. Ereignis

(Zahl-Adler) kann nämlich in zweifacher Weise zustande kommen: Das 1. Mal Zahl und das 2. Mal Adler oder umgekehrt das 1. Mal Adler und das 2. Mal Zahl. Richtig wäre folgende Überlegung gewesen: Es resultieren 4 gleichwertige Ereignisse: Zahl-Zahl, Adler-Adler, Zahl-Adler und Adler-Zahl. Daraus ersehen wir, dass die Wahrscheinlichkeit, 2-mal Zahl zu werfen, nicht $p = 1/3$, sondern $p = 1/4$ ausmacht. Dadurch, dass wir die Aufeinanderfolge von Zahl und Adler außer Acht gelassen haben, sind die Ereignisse nicht mehr gleich wahrscheinlich bzw. nicht mehr gleichwertig.

1.1.1 Theoretische und empirische Wahrscheinlichkeit

Wenn wir eine Münze werfen, so erwarten wir das Resultat „Zahl“ mit einer Wahrscheinlichkeit von $p = 1/2$. Wir folgern nämlich: Es gibt nur 2 mögliche Resultate, von denen eines im gegebenen Fall mit Sicherheit eintreten muss, so dass – wenn die Münze nicht verfälscht ist – jedes der beiden Resultate die gleiche Wahrscheinlichkeit hat. Da wir dieses Resultat allein auf logischem Weg erzielt haben, sprechen wir von einer theoretischen, einer erwarteten oder einer *A-priori-Wahrscheinlichkeit*.

Werfen wir dagegen eine Münze, deren eine Kante stark abgenutzt ist, so dürfen wir nicht mehr erwarten, dass bei einem beliebigen Wurf das Symbol „Zahl“ mit der Wahrscheinlichkeit $p = 1/2$ oben liegt. Auf die Größe der Wahrscheinlichkeit, in diesem Fall „Zahl“ zu werfen, kann uns nur ein Experiment einen Hinweis geben: Wir werfen die Münze einige 100-mal und zählen aus, wie oft wir das Resultat „Zahl“ erhalten. Bilden wir den Quotienten aus der Anzahl der „Zahlen“ und der Anzahl der Würfe, so erhalten wir eine relative Häufigkeit, die wir als empirische, beobachtete oder als *A-posteriori-Wahrscheinlichkeit* bezeichnen. Mit zunehmender Anzahl von Versuchen konvergiert die relative Häufigkeit auf einen konstanten Wert p. Bezeichnen wir die Häufigkeit eines Ereignisses A mit $f(A)$ und die Anzahl aller Ereignisse einer Versuchsreihe mit N, so ergibt sich als Gleichung für die A-posteriori-Wahrscheinlichkeit $p(A)$:

$$p(A) = \lim_{N \to \infty} \frac{f(A)}{N} \tag{1.2}$$

Im Folgenden wenden wir uns den wichtigsten Gesetzen der Wahrscheinlichkeitsrechnung zu, dem Additions- und Multiplikationssatz für Wahrscheinlichkeiten.

1.1.2 Additions- und Multiplikationssatz

Beim Würfelspiel können wir uns fragen, wie groß die Wahrscheinlichkeit ist, eine 6 *oder* eine 5 zu werfen. Da wir es hier mit 2 günstigen unter 6 möglichen Fällen zu tun haben, ist $p = 2/6 = 0{,}33$. Die Wahrscheinlichkeit, eine 6, eine 5 oder eine 2 zu werfen, ist entsprechend durch 1/6 + 1/6 + 1/6 = 0,5 gegeben. Sie ist also die Summe der Wahrscheinlichkeiten, eine 6, eine 5 oder eine 2 zu werfen. Die Verallgemeinerung dieser Überlegung führt zum *Additionssatz der Wahrscheinlichkeit.* Er lautet:

> Die Wahrscheinlichkeit p, dass von k einander ausschließenden Ereignissen das erste *oder* das zweite *oder* das dritte *oder* das k-te eintritt, ist gleich der Summe der Wahrscheinlichkeiten für das Auftreten der k Einzelereignisse.

Bezeichnen wir allgemein mit p_i die Wahrscheinlichkeit des i-ten Ereignisses, so beträgt die zusammengesetzte Wahrscheinlichkeit nach dem Additionssatz:

$$p = p_1 + p_2 + \ldots + p_i + \ldots + p_k = \sum_{i=1}^{k} p_i \qquad (1.3)$$

Wenn k die Anzahl *aller* möglichen Ereignisse eines Versuchs (z. B. die Augenzahlen 1 bis 6 bei einem Würfelversuch) kennzeichnet, hat die Gesamtsumme der Einzelwahrscheinlichkeiten den Wert 1 (die Wahrscheinlichkeit eine 1 *oder* eine 2 *oder* eine …6 zu werfen, ist 1. Anders formuliert: Die Wahrscheinlichkeit, dass bei einem Würfelversuch irgendeine der 6 möglichen Augenzahlen fällt, beträgt 1).

Nun zu einem anderen Problem:

Wenn wir einen Würfel 2-mal hintereinander werfen, so können wir uns fragen: Wie groß ist die Wahrscheinlichkeit p, dass wir 2-mal eine 6, d. h. also eine 6 *und* eine 6 werfen? Dieselbe Frage wäre auch für den gleichzeitigen Wurf zweier Würfel zu stellen. Die theoretische Wahrscheinlichkeit leitet sich aus folgender Überlegung her: Für den 1. Wurf gibt es 6 mögliche Ereignisse, nämlich die Zahlen 1 bis 6. Das gleiche gilt für den 2. Wurf. Da nun das Ereignis des 2. Wurfs vom Ereignis des 1. Wurfs unabhängig ist, treten alle möglichen Ereigniskombinationen (1 und 1, 1 und 2, 1 und 3, …, 6 und 5, 6 und 6) mit gleicher Wahrscheinlichkeit auf. Da jede mögliche Zahl des 1. Wurfs mit jeder möglichen Zahl des 2. Wurfs kombiniert sein kann, erhält man $6 \cdot 6 = 36$ gleichwertige mögliche Ereignisse. Eines dieser Ereignisse, nämlich die Kombination 6 und 6, stellt das günstige Ereignis dar, so dass wir nach Gl. 1.1 den Wert $p = 1/36$ errechnen.

Entsprechend ist die Wahrscheinlichkeit, mit einer Münze 2-mal „Zahl“ zu werfen: $p = 1/2 \cdot 1/2 = 1/4$. Wir können diesen als *Multiplikationssatz der Wahrscheinlichkeit* bekannten Tatbestand allgemein so formulieren:

> Die Wahrscheinlichkeit p, dass von k unabhängigen Ereignissen das erste *und* das zweite *und* das dritte *und* …das k-te Ereignis gemeinsam auftreten, ist gleich dem Produkt der Einzelwahrscheinlichkeiten p_i dieser Ereignisse.

$$p = p_1 \cdot p_2 \cdot \ldots \cdot p_i \cdot \ldots \cdot p_k = \prod_{i=1}^{k} p_i \tag{1.4}$$

Mit einem einfachen „Trick" lässt sich schnell erkennen, wann der Additionssatz und wann der Multiplikationssatz anzuwenden ist. Wenn die Ereignisse mit einer „oder"-Verknüpfung zu verbinden sind, kommt der Additionssatz zur Anwendung und bei einer „und"-Verknüpfung der Multiplikationssatz.

Frage: Wie groß ist die Wahrscheinlichkeit (p), mit zwei Würfeln eine Augenzahlsumme von 10 zu werfen? Wir erzielen diese Summe mit einer 5 für den ersten *und* einer 5 für den zweiten Würfel *oder* einer 4 für den ersten *und* einer 6 für den zweiten Würfel *oder* mit einer 6 für den ersten *und* einer 4 für den zweiten Würfel. Wir wenden bei den „und"-Verknüpfungen den Multiplikationssatz an und bei den „oder"-Verknüpfungen den Additionssatz.

$$\begin{array}{lccccccccc} & 5 & \text{und} & 5 & \text{oder} & 4 & \text{und} & 6 & \text{oder} & 6 \ \text{und} \ 4 \\ p = & 1/6 & \cdot & 1/6 & + & 1/6 & \cdot & 1/6 & + & 1/6 \cdot 1/6 = 3/36 \end{array}$$

Ein anderes Beispiel: Die Augenzahlsumme 7 resultiert aus 6 und 1 oder 5 und 2 oder 4 und 3 oder 3 und 4 oder 2 und 5 oder 1 und 6, d.h. $p = 6/36 = 1/6$.

Additions- und Multiplikationssatz sind wichtige Ausgangspunkte der folgenden Ausführungen und der späteren über die statistische Entscheidung (▶ Abschn. 1.2.3).

1.1.3 Punktwahrscheinlichkeit

Wenden wir uns von den Würfelversuchen, die 6 mögliche Resultate ergeben, wieder dem einfacheren Münzenversuch mit 2 Alternativen zu: Fragen wir uns, welche Kombinationen von „Zahl" (Z) und „Adler" (A) wir bei einem Wurf von 3 Münzen theoretisch erhalten können. Im Folgenden sind die Möglichkeiten vollzählig zusammengestellt: ZZZ, ZZA, ZAZ, AZZ, ZAA, AZA, AAZ, AAA.

Es gibt also 8 mögliche Kombinationen, nämlich einmal 3Z, 3-mal 2Z und 1A, 3-mal 1Z und 2A und einmal 3A. Der 1. Münzwurf hat 2 mögliche Ausgänge (Z, A) und der 2. auch. Kombinieren wir die beiden Münzwürfe, resultieren $2 \cdot 2 = 4$ mögliche Kombinationen, denn jeder Ausgang des 1. Münzwurfs kann mit jedem Ausgang des 2. Münzwurfs gemeinsam auftreten (ZZ, ZA, AZ, AA). Kommt nun noch eine 3. Münze hinzu, kann jede dieser 4 Kombinationen mit den beiden Ausgängen des 3. Münzwurfs kombiniert werden, d. h. wir erhalten die oben schon ausgeführten $2 \cdot 2 \cdot 2 = 2^3 = 8$ verschiedenen Kombinationen. Allgemein erhalten wir bei N Münzen 2^N mögliche Kombinationen.

Bei $N = 3$ Münzen finden wir unter den 8 möglichen Kombinationen nur eine, bei der alle Münzen auf „Zahl" fallen. Die Wahrscheinlichkeit, 3-mal „Zahl" zu erhalten, ist also nach Gl. 1.1 $p = 1/8$. Die Wahrscheinlichkeit, dass wir bei einem Wurf die Kombination 2-mal „Zahl" und einmal „Adler" (ZZA *oder* ZAZ *oder* AZZ) antreffen werden, beträgt 3/8 wie auch für die Kombination einmal „Zahl“ und 2-mal „Adler“ (ZAA *oder* AZA *oder* AAZ). Die Wahrscheinlichkeit, 3-mal „Adler“ zu werfen, ergibt sich wiederum zu 1/8.

> Die Wahrscheinlichkeit für ein bestimmtes Ereignis (z. B. die Kombination „2-mal Z, einmal A“) nennt man *Punktwahrscheinlichkeit*. Man erhält die Punktwahrscheinlichkeit p, indem man die Häufigkeit, mit der das Ereignis vorkommen kann (z. B. 3 Möglichkeiten für die Kombination „2-mal Z und einmal A“), durch die Anzahl aller möglichen Kombinationen (in diesem Beispiel 8) dividiert.

Diese p-Werte erhalten wir auch über das sog. *Pascal-Dreieck* (Tab. 1.1). Das Pascal-Dreieck in Tabelle 1.1 wurde für $N = 1$ bis $N = 5$ Münzen in Einserschritten entwickelt. (Die in Klammern gesetzte Zeile $N = 0$ wurde der Vollständigkeit halber mit aufgenommen). Wie man leicht erkennt, ergeben sich die Häufigkeiten einer Zeile als Summe von jeweils 2 benachbarten Häufigkeiten der vorangehenden Zeile. Jede Zeile wird am Anfang und am Ende durch die Zahl „1“ ergänzt. Diesem Prinzip folgend lässt sich das Pascalsche Dreieck in Tabelle 1.1 beliebig fortschreiben.

Die Spalte N enthält die Anzahl der Münzen und die Spalte 2^N die Anzahl *möglicher* Kombinationen. Die Zahlen innerhalb des Dreiecks geben an, wie häufig die jeweilige Kombination vorkommt („*günstige*“ Kombinationen). Die 3. Zeile (mit $N = 2$) als Beispiel besagt also, dass die Kombination „$2 \times$Z“ einmal, die Kombination „$1 \times$Z, $1 \times$A“ 2-mal und die Kombination „$2 \times$A“ einmal vorkommt, so dass insgesamt $2^2 = 4$ mögliche Kombinationen resultieren.

Nach der Regel „günstige Kombinationen/mögliche Kombinationen“ (Gl. 1.1) ergeben sich aus Tabelle 1.1 z. B. für $N = 4$ die folgenden Punktwahrscheinlichkeiten:

p (4-mal Z) = 1/16
p (3-mal Z, einmal A) = 4/16
p (2-mal Z, 2-mal A) = 6/16
p (einmal Z, 3-mal A) = 4/16
p (4-mal A) = 1/16

Die Punktwahrscheinlichkeiten für alle Varianten „günstiger“ Kombinationen addieren sich zu 1.

Diese Punktwahrscheinlichkeiten lassen sich jedoch auch ohne das Pascal-Dreieck ermitteln. Da wir die Anzahl der möglichen Kombinationen mit 2^N bereits kennen, brauchen wir uns nur noch darüber Gedanken zu machen, wie häufig eine günstige Kombination zustandekommen kann, um über Gl. 1.1 deren Punktwahrscheinlichkeit ausrechnen zu können.

Angenommen, wir hätten nacheinander 4 Münzen geworfen und fragen danach, wie häufig die „günstige“ Kombination „$3 \times$Z, $1 \times$A“ vorkommen kann. Nach Tabelle 1.1 gibt es hierfür offenbar 4 verschiedene Möglichkeiten, d.h. die Anzahl der günstigen Möglichkeiten wäre 4.

Zu diesem Ergebnis kommen wir auch durch folgende Überlegung: Wir markieren zunächst 4 Münzen mit den Buchstaben A, B, C und D. Nun werfen wir die 4 Münzen in einer zufälligen Reihenfolge und fragen, mit welcher Münze die 1. „Zahl“ geworfen wird. Dies kann natürlich mit gleicher Wahrscheinlichkeit jede der 4 Münzen sein. Wenn wir annehmen, dass mit A die 1. „Zahl“ geworfen wurde, bleiben für die 2. „Zahl“ nur die Münzen B, C und D übrig. Hätten wir mit Münze B die 2. Zahl geworfen, kann die 3. Zahl nur noch auf die Münzen C oder D fallen. Fällt sie auf C, bleibt für „Adler“ nur die Münze D übrig. Insgesamt gibt es also $4 \cdot 3 \cdot 2 \cdot 1 = 24$ Möglichkeiten, das Ergebnis „3×Z, 1×A“ zu erzielen. Diese sind in Tabelle 1.2 veranschaulicht.

Betrachten wir diese 24 Aufteilungsmöglichkeiten, stellen wir fest, dass es z.B. für die Aufteilung von Z auf die Münzen A, B und C 6 verschiedene Reihenfolgen gibt. Diese sind in Tabelle 1.2 durch ein Kreuz markiert. Die 6

Tabelle 1.1. Pascal-Dreieck

Überwiegen von „Zahl“ … Überwiegen von „Adler“	N	2^N
(1)	(0)	(1)
1 1		
1×Z 1×A	1	2
1 2 1		
2×Z 1×Z, 1×A 2×A	2	4
1 3 3 1		
3×Z 2×Z, 1×A 1×Z, 2×A 3×A	3	8
1 4 6 4 1		
4×Z 3×Z, 1×A 2×Z, 2×A 1×Z, 3×A 4×A	4	16
1 5 10 10 5 1		
5×Z 4×Z, 1×A 3×Z, 2×A 2×Z, 3×A 1×Z, 4×A 5×A	5	32

Tabelle 1.2. Verteilungsmöglichkeiten für die Kombination „3×Z, 1×A“ auf 4 Münzen A, B, C und D. (Erläuterungen s. Text)

Nr.	Z	Z	Z	A	Nr.	Z	Z	Z	A
X 1	A	B	C	D	X 13	C	A	B	D
2	A	B	D	C	14	C	A	D	B
X 3	A	C	B	D	X 15	C	B	A	D
4	A	C	D	B	16	C	B	D	A
5	A	D	B	C	17	C	D	A	B
6	A	D	C	B	18	C	D	B	A
X 7	B	A	C	D	19	D	A	B	C
8	B	A	D	C	20	D	A	C	B
X 9	B	C	A	D	21	D	B	A	C
10	B	C	D	A	22	D	B	C	A
11	B	D	A	C	23	D	C	A	B
12	B	D	C	A	24	D	C	B	A

Reihenfolgen ergeben sich nach der gleichen Zählregel wie die oben ermittelten 24 Aufteilungsmöglichkeiten. Bei 3 Münzen gibt es für den 1. Platz 3 Möglichkeiten für Z, für den 2. Platz gibt es 2 Möglichkeiten und die dritte Münze erhält den letzten Platz. Insgesamt sind also $3 \cdot 2 \cdot 1 = 6$ Reihenfolgen möglich. Hierbei handelt es sich um die Reihenfolgen ABC, ACB, BAC, BCA, CAB und CBA. Für die Frage nach der Anzahl der günstigen Kombinationen für „3×Z, 1×A" sind diese 6 Reihenfolgen jedoch irrelevant, denn alle bedeuten 3-mal Z für die Münzen A, B und C.

Andere günstige Kombinationen resultieren, wenn Z auf die Münzen A, B und D; A, C und D oder B, C und D fällt. Jede dieser günstigen Kombinationen kann ihrerseits in 6 verschiedenen, vom Ergebnis her gleichwertigen Reihenfolgen vorkommen. Um die Anzahl der günstigen Kombinationen für „3×Z, 1×A" zu erhalten, dividieren wir also alle 24 Aufteilungsmöglichkeiten durch die Anzahl der jeweils 6 „ergebnisneutralen" Reihenfolgen. Im Resultat erhalten wir mit $24:6=4$ die bereits bekannte Anzahl für die Kombinationen „3×Z, 1×A".

Ein anderes Beispiel: Wieviele Möglichkeiten gibt es für die Kombination „2×Z, 2×A" bei 4 Münzen? Die Antwort liefert uns das Pascal-Dreieck: 6 Möglichkeiten.

Der folgende Gedankengang bestätigt diese Antwort. Wie oben gibt es 24 Möglichkeiten, wie sich die Kombination „2×Z, 2×A" auf die Münzen A, B, C und D verteilen kann: Die erste Zahl fällt auf eine der 4 Münzen, die zweite auf eine der verbleibenden 3 Münzen, der 1. Adler auf eine der 2 restlichen Münzen und der 2. Adler auf die letzte noch „freie" Münze. Dies gibt erneut $4 \cdot 3 \cdot 2 \cdot 1 = 24$ Verteilungsmöglichkeiten. Hierin enthalten sind jedoch auch diesmal verschiedene, im Ergebnis neutrale Abfolgen. Für „2×Z" auf den Münzen A, B gibt es 2 ergebnisneutrale Abfolgen (AB, BA) und für „2×A" auf C, D ebenfalls (CD, DC). Dies gilt entsprechend für alle anderen Kombinationen, so dass wir die Anzahl aller Verteilungsmöglichkeiten durch die Anzahl der jeweils ergebnisneutralen Abfolgen (2 für „2×Z" und 2 für „2×A") dividieren, um die Anzahl aller Möglichkeiten für die günstige Kombination „2Z, 2A" zu erhalten. Das Ergebnis lautet – wie bereits bekannt – $24/(2 \cdot 2) = 6$: (ZZAA; ZAZA; ZAAZ; AZZA; AZAZ; AAZZ).

Für die Verallgemeinerung dieses Gedankenganges vereinbaren wir N = Anzahl der Münzen, x = Häufigkeit für „Zahl" und demzufolge N–x = Häufigkeit für „Adler". Die Anzahl der Verteilungsmöglichkeiten für eine beliebige Kombination von Z und A bei N Münzen ergibt sich zu

$$N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot 2 \cdot 1 = N!$$

Die obige Produktkette wird durch N! (lies: N Fakultät) abgekürzt. N! ist zu dividieren durch die Anzahl der ergebnisneutralen Abfolgen für x-mal Zahl (x!) und für (N–x)-mal Adler (N–x)! Wir erhalten also mit M = Anzahl der Möglichkeiten für eine beliebige Kombination

$$M = \frac{N!}{x! \cdot (N-x)!}$$

Dieser Ausdruck wird häufig durch das sog. *Euler-Symbol* $\binom{N}{x}$ (lies: N über x) gekennzeichnet.

$$M = \frac{N!}{x! \cdot (N-x)!} = \binom{N}{x} \qquad (1.5)$$

Mit dieser Rechenregel können wir also auch ohne das Pascal-Dreieck für beliebige N- und x-Werte die Anzahl der Möglichkeiten für eine Kombination (=günstige Ereignisse) ermitteln. Für N=5 und x=3 ergibt sich beispielsweise

$$\binom{5}{3} = \frac{5!}{3! \cdot (5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1) \cdot (2 \cdot 1)} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

Dieser Wert stimmt mit dem entsprechenden Wert des Pascal-Dreiecks überein (10 Möglichkeiten für x=3-mal Z bzw. für „3×Z, 2×A" bei N=5). Die Punktwahrscheinlichkeit, mit 5 Münzwürfen genau 3-mal Zahl zu erzielen, ergibt sich (wegen $2^5 = 32$ = Anzahl der möglichen Kombinationen) über Gl. 1.1 also zu p=10/32= 0,31.

In allgemeiner Schreibweise errechnen wir eine Punktwahrscheinlichkeit im Münzwurfbeispiel nach folgender Beziehung

$$p(x) = \frac{\binom{N}{x}}{2^N} = \binom{N}{x} \cdot \left(\frac{1}{2}\right)^N \qquad (1.6)$$

Die Punktwahrscheinlichkeiten für unterschiedliche x- und N-Werte konstituieren eine Verteilung, die als *Binomialverteilung* für gleich wahrscheinliche Alternativereignisse bezeichnet wird. Wie diese Verteilung praktisch genutzt werden kann, werden wir im ► Abschn. 2.1.1 zeigen.

Zu Gl. 1.5 sind noch einige Anmerkungen erforderlich. Setzen wir N=x, ergibt sich N!/N!·0! Mit 0!=1 ergibt sich $\binom{N}{N} = 1$. Begründung: N!=(N–1)!·N oder (N–1)!=N!/N. Hieraus folgt für N=1: 0!=1!/1=1. Außerdem ist Gl. 1.5 zu entnehmen, dass $\binom{N}{x} = \binom{N}{N-x}$ sein muss. Bezogen auf das Münzwurfbeispiel besagt diese Äquivalenz, dass die Häufigkeit der Kombination x-mal „Zahl" bei N Würfen mit der Häufigkeit der Kombination (N–x)-mal „Zahl" bei N Würfen übereinstimmt, also z. B. $\binom{5}{3} = \binom{5}{2} = 10$.

1.1.4 Überschreitungswahrscheinlichkeit

Wir werden im Folgenden noch eine andere Wahrscheinlichkeit kennenlernen, die sich am besten anhand eines Wettbeispiels einführen lässt: Angenommen, wir haben gewettet, mit N=4 Münzen mindestens x=3-mal „Zahl" zu werfen. Wie groß ist die Wahrscheinlichkeit, diese Wette zu gewinnen? Die Antwort ist einfach: „Mindestens 3-mal" bedeutet, 3-mal oder 4-mal „Zahl" zu werfen; also ist die gesuchte Wahrscheinlichkeit – wir bezeichnen sie mit (groß) P und nennen sie *Überschreitungswahrscheinlichkeit* – nach dem Additionssatz gleich der

Punktwahrscheinlichkeit, 3-mal „Zahl“ zu werfen: $p(x=3)=4/16$ plus der Punktwahrscheinlichkeit, 4-mal „Zahl“ zu werfen: $p(x=4)=1/16$; also ist $P=4/16+1/16=5/16$. In gleicher Weise könnten wir nach der Wahrscheinlichkeit, mindestens 2-mal „Zahl“ zu werfen, fragen. Sie beträgt für $x=2$, $x=3$ und $x=4$ $P=6/16+4/16+1/16=11/16$.

Wir können die Überschreitungswahrscheinlichkeit definieren als die Wahrscheinlichkeit des Auftretens eines bestimmten Ereignisses, vermehrt um die Wahrscheinlichkeiten aller „extremeren“ Ereignisse.

Statt nach der Wahrscheinlichkeit für „mindestens 3-mal Zahl“ hätten wir auch nach der Wahrscheinlichkeit für „höchstens einmal Adler“ fragen können. Für beide Fälle ist die Überschreitungswahrscheinlichkeit natürlich identisch.

Allgemein: Die Wahrscheinlichkeit, dass ein Ereignis A bei N Versuchen mindestens x-mal auftritt, entspricht der Wahrscheinlichkeit, dass das zu A komplementäre Ereignis höchstens (N–x)-mal auftritt.

Unter Verwendung der Gl. 1.6 ergibt sich die Überschreitungswahrscheinlichkeit dafür, dass x mindestens einen Wert k annimmt, in folgender Weise:

$$P(x \geq k) = \left(\frac{1}{2}\right)^N \cdot \sum_{x=k}^{N} \binom{N}{x} \tag{1.7}$$

Fragen wir nach der Überschreitungswahrscheinlichkeit, mit der x höchstens so groß wie ein beliebiger Wert k ist, errechnet man

$$P(x \leq k) = \left(\frac{1}{2}\right)^N \cdot \sum_{x=0}^{k} \binom{N}{x} \tag{1.8}$$

Beide Gleichungen sind äquivalent, wenn man eine Gleichung auf das Ereignis A (z. B. Zahl) und die andere auf das Komplementärereignis (z. B. Adler) anwendet. So entspricht z. B. die Überschreitungswahrscheinlichkeit, bei $N=10$ Münzwürfen mindestens 8-mal Zahl zu werfen ($k=8$ in Gl. 1.7), der Überschreitungswahrscheinlichkeit für höchstens 2-mal Adler ($k=2$ in Gl. 1.8). In beiden Fällen erhält man $P=0{,}055$.

1.1.5 Einseitige und zweiseitige Überschreitungswahrscheinlichkeit

Im obigen Beispiel (mindestens $x=8$-mal Zahl bei $N=10$ Münzwürfen) haben wir eine einseitige Überschreitungswahrscheinlichkeit bestimmt. Was unter einer zweiseitigen Überschreitungswahrscheinlichkeit zu verstehen ist, lässt sich wie folgt illustrieren: Wir wetten, bei 4 Würfen entweder 4-mal oder keinmal „Zahl“ zu werfen. Wie groß ist die Chance, diese Wette zu gewinnen? Die Punktwahrscheinlichkeit für $x=4$ beträgt $p(x=4)=1/16$ und die Punktwahrscheinlichkeit für $x=0$ ist $p(x=0)=1/16$, so dass die zweiseitige Überschreitungswahrscheinlichkeit, die wir durch P' kennzeichnen, mit $P'=2/16$ der doppelten einseitigen Überschreitungswahrscheinlichkeit entspricht.

Hätten wir gewettet, mit 4 Münzen mindestens 3-mal „Zahl“ oder höchstens einmal „Zahl“ zu werfen, so wäre dies ebenfalls eine zweiseitige Wette, deren Gewinnchance nach dem Pascal-Dreieck oder über Gl. 1.6 wie folgt zu berechnen wäre: Mindestens 3-mal „Zahl“ heißt 3- oder 4-mal „Zahl“, deren Punktwahrscheinlichkeiten 4/16 und 1/16 betragen. Hinzu kommen die Punktwahrscheinlichkeiten für einmal „Zahl“ ($p=4/16$) und für keinmal „Zahl“ ($p=1/16$). Die gesamte zweiseitige Überschreitungswahrscheinlichkeit beträgt also $P'=1/16+4/16+4/16+1/16=10/16$.

Die Frage, ob es sich um eine einseitige oder zweiseitige Wette oder – in der Terminologie der Statistik – um einen einseitigen oder zweiseitigen Test handelt, ist für die Überprüfung bestimmter empirischer Fragestellungen von großer Bedeutung. (Beispiel: Die Überprüfung der Frage, ob eine neue Behandlungsmethode einer herkömmlichen Methode überlegen ist, erfordert einen einseitigen Test, während der Nachweis eines ungerichteten Unterschiedes – die neue Methode ist besser *oder* schlechter als die herkömmliche – über einen zweiseitigen Test zu führen wäre.) Wir werden darauf später (▶ Abschn. 1.2.3) noch zurückkommen.

Festzuhalten ist, dass die Wahrscheinlichkeit für den zweiseitigen Test durch Verdopplung der Wahrscheinlichkeit für den einseitigen Test zu ermitteln ist, sofern – wie im Münzwurfbeispiel – die Wahrscheinlichkeitsverteilung für x symmetrisch ist.

1.2 Statistische Hypothesenprüfung

Wie bereits eingangs gesagt, befassen wir uns in diesem Buch primär mit der statistischen Überprüfung klinisch-wissenschaftlicher Hypothesen. Hierfür sind häufig die verteilungsfreien oder nonparametrischen Verfahren besser geeignet als die „klassischen“ parametrischen Verfahren. Die Anwendung parametrischer Verfahren bereitet insbesondere bei der Untersuchung kleinerer Stichproben Probleme, wenn wichtige Voraussetzungen – z. B. normalverteilte Merkmale – verletzt sind (▶ Abschn. 1.2.6). Bevor wir auf die verteilungsfreien Verfahren ausführlich eingehen, sollen in einem Überblick die wichtigsten Stationen der statistischen Hypothesenprüfung – die Überführung von inhaltlichen Forschungshypothesen in statistische Hypothesen, das Konzept der statistischen Signifikanz und der klinisch-praktischen Bedeutsamkeit sowie die Spezifika verteilungsfreier Tests – vorgestellt werden.

Von besonderer Bedeutung, nicht nur für hypothesenprüfende Untersuchungen, ist eine ausführliche Versuchsplanung, auf die wir zunächst eingehen.

1.2.1 Versuchsplanung

Es ist verständlich, wenn junge, ambitionierte Doktoranden nach Festlegung ihrer Fragestellung möglichst rasch das benötigte klinische Datenmaterial erheben und auch auswerten wollen; hiervor sei jedoch mit Nachdruck gewarnt, denn das Gelingen und die Aussagekraft einer Studie hängen zu einem erheblichen Teil von einer sorgfältigen, ausführlichen Versuchsplanung ab, die vor der eigentlichen Datenerhebung durchzuführen ist. Erfahrene Empiriker sind der Auffassung, dass mindestens 50% der gesamten Arbeitszeit für eine Studie auf die Versuchsplanung entfallen sollten. Außerdem wird dringend empfohlen, sich schon im Planungsstadium mit einem Bio- oder Medizinstatistiker abzusprechen und nicht erst – wie so oft – nachdem die Versuchsergebnisse bereits vorliegen.

Die Ergebnisse der Versuchsplanung sind in einem sog. Studienprotokoll zusammenzufassen. Zum *Studienprotokoll* gehören v. a. folgende Angaben:

Inhalte eines Studienprotokolls

- Eine präzise Formulierung der Fragestellung bzw. ggf. der zu prüfenden Forschungshypothese (Einflussgrößen, Zielgrößen, Störgrößen).
- Festlegung und Begründung des Studientyps (Doppelblindversuch? Epidemiologische Studie? Klinisches Experiment?).
- Auswahl und Rekrutierung der Patientenstichprobe (Kontroll- und Experimentalgruppe? Ein- und Ausschlusskriterien für die Patientenauswahl? „Matched samples“? Größe der Stichprobe?). Entsprechende Angaben sind auch für Tierversuche erforderlich.
- Behandlung der Patienten (Welcher Patient erhält welche Behandlung? Zeitliche Abfolge von Kontrolluntersuchungen? Vorgehen bei Komplikationen?).
- Operationalisierung der untersuchungsrelevanten Merkmale (Labortests? Expertenratings? Indikatoren für den Behandlungserfolg?).
- Angaben zur Untersuchungsdurchführung (Klinisches Hilfspersonal? Ausreichende Labortechnik? Belastung der Patienten?).
- Planung der statistischen Auswertung (Welcher Signifikanztest? Risiko I. und ggf. auch II. Art?).

Im Folgenden wollen wir diese Inhalte des Studienprotokolls ausführlicher erläutern [zu den Besonderheiten eines Studienprotokolls bei der Überprüfung von Arzneimitteln vgl. Feiden (1988) bzw. ► S. 22 f].