



Forschungsmethoden und Statistik

für Psychologen und Sozialwissenschaftler

4., aktualisierte und erweiterte Auflage

Peter Sedlmeier
Frank Renkewitz



Pearson

Der Begriff Freiheitsgrade steht für die Anzahl der Werte, die in einem statistischen Ausdruck frei variieren können. Ein solcher statistischer Ausdruck ist beispielsweise die Stichprobenvarianz. Bei der Berechnung von s^2 können von den n Werten nur $n - 1$ frei variieren, da die Summe aller Abweichungen der n Einzelwerte von ihrem Mittelwert immer null ist (siehe ▶ Kapitel 6):

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

Wenn also z.B. $n = 4$ und die ersten drei Abweichungen vom Mittelwert 6, -9 und -1 betragen, muss die vierte Abweichung $0 - 6 + 9 + 1 = 4$ sein. Wenn wir zwei Mittelwerte für einen statistischen Ausdruck berechnen würden, dann wären die Freiheitsgrade $n - 2$. Als Faustregel gilt, dass man für jeden Mittelwert, der für die Berechnung eines statistischen Ausdrucks benötigt wird, einen Freiheitsgrad abziehen muss.

In unserem Beispiel müssen wir zwei Mittelwerte berechnen, einen für die Jungen und einen für die Mädchen. Wir haben insgesamt $n = 10$ Werte und die Freiheitsgrade betragen somit $n - 2 = 8$. Sie erinnern sich: Unsere Ergebnisse waren $d = 1,06$ und $g = 0,95$. Eingesetzt in die jeweiligen Formeln erhalten wir diese Werte wieder (gerundet):

$$d = g\sqrt{\frac{n}{df}} = 0,95\sqrt{\frac{10}{8}} = 1,06$$

und

$$g = d\sqrt{\frac{df}{n}} = 1,06\sqrt{\frac{8}{10}} = 0,95$$

9.4.2 Korrelationen aus Abstandsmaßen

Wenn in einer Studie die Effektgrößen als Abstandsmaße angegeben wurden, man aber gerne Korrelationen hätte, kann man folgende Formeln verwenden, sofern die zwei Stichprobengrößen gleich sind:⁶

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

und

$$r = \frac{g}{\sqrt{g^2 + 4\left(\frac{df}{n}\right)}}$$

Bei ungleichen Stichprobengrößen ist die Sache etwas komplizierter:

$$r = \frac{d}{\sqrt{d^2 + \frac{1}{pq}}}$$

wobei p und q die Anteile der jeweiligen Stichprobengrößen an der Gesamtstichprobengröße sind. Wenn in der ersten Gruppe beispielsweise $n_A = 40$ und in der zweiten $n_B = 60$ Personen untersucht worden wären, dann erhielten wir $p = 0,4$ und $q = 0,6$. Wenn beide Anteile 0,5 sind, wenn also die Gruppengrößen gleich sind, dann erhält man $1/pq = 4$. Die entsprechende Umrechnungsformel für g ist:

$$r = \sqrt{\frac{g^2(n_A n_B)}{g^2(n_A n_B) + (n_A + n_B)df}}$$

6 Im Einstichprobenfall oder bei Differenzwerten macht eine solche Umrechnung keinen Sinn.

Für unser Beispiel ergäben sich wieder folgende Werte (wir benutzen zu Demonstrationszwecken die komplizierteren, aber allgemeineren Formeln):

$$r = \frac{d}{\sqrt{d^2 + \frac{1}{pq}}} = \frac{1,06}{\sqrt{1,06^2 + \frac{1}{0,5 \cdot 0,5}}} = 0,47$$

und

$$r = \sqrt{\frac{g^2(n_A n_B)}{g^2(n_A n_B) + (n_A + n_B)df}} = \sqrt{\frac{0,95^2(5 \cdot 5)}{0,95^2(5 + 5) + (5 + 5)8}} = 0,47$$

9.4.3 Abstandsmaße aus Korrelationen

Man kann natürlich auch Abstandsmaße aus Korrelationen berechnen. Die entsprechenden Formeln (allgemeiner Fall) und zugleich auch die Ergebnisse für unser Beispiel sind:

$$d = \frac{\sqrt{\frac{1}{pq}} \cdot r}{\sqrt{1 - r^2}} = \frac{\sqrt{\frac{1}{0,5 \cdot 0,5}} \cdot 0,47}{\sqrt{1 - 0,47^2}} = 1,06$$

und

$$g = \frac{r}{\sqrt{1 - r^2}} \sqrt{\frac{(n_A + n_B)df}{n_A n_B}} = \frac{0,47}{\sqrt{1 - 0,47^2}} \sqrt{\frac{(5 + 5)8}{5 \cdot 5}} = 0,95$$

9.5 Wie bedeutsam ist eine Effektgröße?

Wir haben beschrieben, wie man Effektgrößen aus den Rohwerten berechnet (siehe auch ▶ Kapitel 7), und wie man sie ineinander überführen kann. Wie aber kann man sie sinnvoll interpretieren? Wann ist ein Effekt bedeutsam? Auf diese Fragen gibt es keine allgemein gültige Antwort. Die am ehesten zutreffende Antwort wäre noch: „Das hängt vom Gegenstandsbereich ab!“ Wenn etwa der Effekt von Psychotherapie bei Angstzuständen im Durchschnitt einem $d = 0,67$ entspricht,⁷ dann ist für eine neu eingeführte Psychotherapieform ein Effekt von $d = 0,3$ wenig interessant, einer von $d = 0,9$ aber sehr wohl. Bei praktischen Maßnahmen wie Trainingsergebnissen oder der Auswirkung von Interventionen weiß man ja – durch die Berechnung von Effektgrößen –, wie groß die bisherigen Effekte waren. Wenn das neue Programm oder die neue Intervention nicht mindestens die gleichen Effekte zeigt, dann ist das häufig ein guter Grund, die bisherige Vorgehensweise nicht zu ändern. Bisweilen sind auch kleine Effekte schon sehr bedeutsam. Das kann man in manchen Fällen durch eine besondere Darstellungsart, den sogenannten *Binomial Effect Size Display (BESD)* verdeutlichen (siehe folgender Kasten).

⁷ Wir werden uns in **Kapitel 28** damit beschäftigen, wie man solche durchschnittlichen Effekte berechnet und interpretiert.

BINOMIAL EFFECT SIZE DISPLAY (BESD) UND KLEINE EFFEKTE

Der BESD wurde von Rosenthal und Rubin (1979) eingeführt und dient zur Illustration einer bestimmten Art von Effektgröße für die Beschreibung von dichotomen oder dichotomisierten Ergebnissen wie z.B. Erfolg vs. Misserfolg, Verbesserung vs. keine Verbesserung, Überlebender vs. Nicht-Überlebender usw. Man kann damit die Korrelation zwischen zwei dichotomen Variablen als *Veränderung in der „Erfolgsquote“* beschreiben. Ein reales Beispiel soll das verdeutlichen (siehe Rosenthal & Rosnow, 1991, 283f.):

In den 80er Jahren nahmen 22.071 Ärzte in den USA an einer Studie teil, in der man herausfinden wollte, ob die regelmäßige Einnahme einer geringen Menge von Aspirin gegen Herzinfarkt vorbeugt. Die eine Hälfte (11.037 Ärzte) erhielt tatsächlich Aspirin während die andere (11.034 Ärzte) ein Placebo verabreicht bekam. Nach fünf Jahren wurde Bilanz gezogen – die Ergebnisse sind in ► Tabelle 9.2 dargestellt.

Bedingung	Kein Herzinfarkt	Herzinfarkt	Insgesamt
Aspirin	10.933(a)	104 (b)	11.037(a + b)
Placebo	10.845(c)	189 (d)	11.034(c + d)
Insgesamt	21.778(a + c)	293(b + d)	22.071

Tabelle 9.2: Ergebnisse einer groß angelegten Studie zur Präventionswirkung von Aspirin gegen Herzinfarkt

Nun kann man eine naheliegende korrelative Effektgröße berechnen, den Phi-Koeffizienten (ϕ), den Sie in ► Kapitel 7 schon kennengelernt haben:

$$\phi = r = \frac{a \cdot d - b \cdot c}{\sqrt{(a + b) \cdot (c + d) \cdot (a + c) \cdot (b + d)}} = \frac{10933 \cdot 189 - 104 \cdot 10845}{\sqrt{11037 \cdot 11034 \cdot 21778 \cdot 293}} = 0,034$$

Das Ergebnis, ein $r = 0,034$, ist offensichtlich ein sehr kleiner Effekt. Trotzdem wurde die Studie nach diesem Ergebnis abgebrochen, weil es die Auftraggeber als unethisch betrachteten, den Ärzten in der Placebo-Gruppe das Aspirin weiterhin vorzuenthalten. Wie kam man zu diesem Schluss? Beim BESD wird die Korrelation üblicherweise so in eine Vierfeldertafel „übersetzt“, dass sich die Zeilen und Spalten zu 100 aufsummieren (► Tabelle 9.3).

Bedingung	Kein Herzinfarkt, %	Herzinfarkt, %	Insgesamt
Aspirin	51,7 (a)	48,3 (b)	100
Placebo	48,3 (c)	51,7 (d)	100
Insgesamt	100	100	200

Tabelle 9.3: BESD-Darstellung der Ergebnisse einer groß angelegten Studie zur Präventionswirkung von Aspirin gegen Herzinfarkt

Überprüfen wir zunächst, ob die ► Tabelle 9.3 tatsächlich eine Korrelation von 0,034 darstellt:

$$\phi = r = \frac{a \cdot d - b \cdot c}{\sqrt{(a + b) \cdot (c + d) \cdot (a + c) \cdot (b + d)}} = \frac{51,7 \cdot 51,7 - 48,3 \cdot 48,3}{\sqrt{100 \cdot 100 \cdot 100 \cdot 100}} = 0,034$$

Wie kommt man auf die Werte in der ► Tabelle 9.3? Man kann sie direkt aus der Korrelation berechnen. Dazu genügt es, einen der zwei unterschiedlichen Werte zu berechnen; der andere ergibt

sich automatisch durch die Einschränkung, dass Zeilen- und Spaltensummen jeweils 100 ergeben müssen. Die Gleichung zur Berechnung des größeren der beiden Werte ist:

$$\text{Größerer Wert} = 50 + \frac{100r}{2}$$

Daraus ergibt sich für unser Beispiel: $50 + (100 \cdot 0,034)/2 = 51,7$, der Wert in den Zellen *a* und *d* in ► Tabelle 9.3. Was ist nun die Aussage, die im BESD enthalten ist? In diesem Fall sagt er aus, dass durch die Gabe von Aspirin die Erfolgsquote (kein Herzinfarkt) um 3,4 Prozentpunkte von 48,3% auf 51,7% gestiegen ist. Wenn man diesen prozentualen Zuwachs in Anzahl geretteter Leben umrechnet, dann ist das ein außerordentlich bedeutsamer Effekt, obwohl er absolut gesehen sehr klein ist. Der BESD ist gut geeignet zur Veranschaulichung von entsprechenden korrelativen Effektgrößen für Personen, die mit einer Korrelation wenig anfangen können, denen aber Prozentwerte vertraut sind (z.B. für Auftraggeber von Evaluationsstudien ohne besondere Methodenkenntnisse).

Wie interpretiert man aber eine Effektgröße, wenn man sie nicht mit bisherigen Ergebnissen vergleichen kann oder wenn man nicht von vornherein weiß, ab wann man einen Effekt als bedeutsam einstufen kann? Wir haben schon in ► Kapitel 7 gesehen, dass Jacob Cohen für diesen Fall Konventionen für die Beurteilung von Korrelationen eingeführt hat. Derartige Konventionen hat er auch für das Abstandsmaß *d* erarbeitet. Cohen hat dies bereits in den 60er Jahren vorgeschlagen, später aber immer wieder leicht modifiziert. Die Werte in ► Tabelle 9.4 stammen aus Cohen (1992). Ursprünglich hatte er die Werte nur anhand von Beispielen begründet. Spätere Überblicksarbeiten (z.B. Sedlmeier & Gigerenzer, 1989) zeigten aber, dass die von Cohen als mittelgroß eingestuften Effekte tatsächlich die sind, die in einem breiten Querschnitt der psychologischen Forschung am häufigsten vorkommen. Diese Konventionen sollten nicht zu ernst genommen werden. Sie sind nichts anderes als ein Notbehelf, wenn man keine andere Art von Information zur Verfügung hat. Außerdem wird offenbar, dass sie auch nicht sehr konsistent sind. Wenn man sie beispielsweise ineinander umrechnet, dann ergibt sich keine genaue Passung zwischen *d* und *r*. Das hängt mit der Natur von Konventionen zusammen: Sie müssen plausibel aussehen. Eine Korrelation von $r = 0,242$, die einem $d = 0,5$ entsprechen würde (bei gleichen Stichprobengrößen), sieht eben nicht so gut aus wie ein $r = 0,3$ (siehe hierzu aber auch ► Kapitel 24).

Größe des Effekts	Art der Effektgröße	
	<i>d</i>	<i>r</i>
klein	0,2	0,1
mittel	0,5	0,3
groß	0,8	0,5

Tabelle 9.4: Konventionen für die Interpretation von Effektgrößen am Beispiel von *d* und *r*

9.6 Weitere Effektgrößen-Maße

Wie schon erwähnt gibt es eine Vielzahl von unterschiedlichen Effektgrößen. Neben Abstands- und Zusammenhangsmaßen werden auch sogenannte *Maße der erklärten Varianz* verwendet, die in manchen Fällen wieder in Korrelationen umgerechnet werden können. Maße der erklärten Varianz kann man auch als quadrierte Zusammenhangsma-

ße im weitesten Sinn betrachten. Das ist gut zu sehen bei einem Maß, das Sie schon aus ▶ Kapitel 8 kennen: beim Determinationskoeffizienten r^2 , der häufig als Effektgröße bei der linearen Regression benutzt wird. (Die Wurzel aus r^2 ist natürlich eine Korrelation.)

Wenn die abhängige Variable in einer Studie intervallskaliert ist, wird man in den meisten Fällen eine der bisher diskutierten Effektgrößen (r , g oder d) nutzen. Manchmal stehen jedoch nur dichotome Ergebnisse zur Verfügung (z.B. „krank“ versus „gesund“, „verbessert“ versus „nicht verbessert“ oder „infiziert“ versus „nicht infiziert“). Wenn man nun jeweils die Ergebnisse für zwei Bedingungen vergleichen möchte (z.B. „Männer versus Frauen“, „Therapie I versus Therapie II“, „am Programm teilgenommen vs. nicht teilgenommen“) gibt es mehrere Möglichkeiten, Effektgrößen für diesen Vergleich zu berechnen. In neuerer Zeit werden, vor allem in medizinischen Bereichen von Psychologie und Sozialwissenschaften vermehrt zwei solche Effektgrößen benutzt, die kurz erläutert werden sollen: *Relative Risiken (RR)* und *Odds Ratios (OR)*.

9.6.1 Relatives Risiko

Das Relative Risiko ist nichts anderes als das Verhältnis des Risikos in zwei Bedingungen. Betrachten wir zur Illustration des Maßes das Beispiel in ▶ Tabelle 9.5, in dem die Anzahl von Personen angegeben wird, die entweder mit *Therapie I* oder *Therapie II* behandelt wurden (die zwei Bedingungen) und bei denen diese Therapien entweder Nebenwirkungen erzeugten oder nicht (das dichotome Ergebnis).

Gruppe	Ergebnis		Keine Nebenwirkung
	Nebenwirkung		
Therapie I	A = 11		B = 120
Therapie II	C = 23		D = 187

Tabelle 9.5: Vierfeldertafel zur Demonstration der Berechnung von *relativen Risiken (RR)* und *Odds Ratios*

Insgesamt 131 Personen (11 + 120) haben sich in diesem Beispiel der Therapie I unterzogen und bei elf davon sind Nebenwirkungen aufgetreten. Das (einfache) Risiko (R) für Nebenwirkung ist also in diesem Fall:

$$R = A/(A + B) = 11/(11 + 120) = 0,084 \text{ oder } 8,4\% \text{ (gerundet)}$$

Wenn man nun die Risiken der zwei Therapieformen miteinander vergleichen möchte, kann man das mithilfe des relativen Risikos (RR) tun:

$$RR = \frac{A/(A + B)}{C/(C + D)}$$

Für das Beispiel in ▶ Tabelle 9.5 ist das relative Risiko für Therapie I im Vergleich zu Therapie II (gerundet):

$$RR = \frac{11/(11 + 120)}{23/(23 + 187)} = 0,77$$

Wenn das Risiko von Nebenwirkungen in beiden Therapien gleich wäre, dann wäre $RR = 1$. Das Ergebnis in unserem Beispiel bedeutet, dass das Risiko dafür, dass Therapie I Nebenwirkungen erzeugt 0,77 mal so groß (also kleiner) ist als bei Therapie II. Therapie I ist also, relativ gesehen, risikoärmer als Therapie II.

9.6.2 Odds Ratio

Ein zweites Maß, das häufig in klinischen Studien verwendet wird, ist die *Odds Ratio*.⁸ Odds werden auch im Alltag häufig benutzt um Wahrscheinlichkeiten und insbesondere Gewinnquoten auszudrücken: „Die Chancen, dass es morgen regnet, stehen 1:10“ oder „Die Chancen, dass die deutsche Mannschaft dieses Spiel gewinnt, stehen 2:1“. Wenn man den Quotienten aus zwei Odds (A/B und C/D) bildet erhält man die Odds Ratio (OR):

$$OR = \frac{A/B}{C/D} = \frac{A \cdot D}{B \cdot C}$$

Sehen wir uns zunächst wieder als Beispiel die Odds dafür an, dass bei Therapie I Nebenwirkungen auftreten (siehe ► Tabelle 9.5):

$$Odds = A/B = 11/120 = 0,092(\text{gerundet})$$

Und nun die *Odds Ratio*, also das Verhältnis der Odds für Therapie I im Vergleich zu Therapie II (Ergebnis gerundet):

$$OR = \frac{A \cdot D}{B \cdot C} = \frac{11 \cdot 187}{23 \cdot 120} = 0,75$$

Auch für die Odds Ratio gilt wieder, dass sie den Wert 1 hat, wenn beide Odds gleich sind. Wenn die Wahrscheinlichkeit für das erste Ereignis (Nebenwirkung bei Therapie I) geringer ist als die für das zweite (Nebenwirkung bei Therapie II), dann ist die Odds Ratio kleiner als 1, im umgekehrten Fall ist sie größer als 1.

Odds Ratios und Relative Risiken können stark voneinander abweichen. Ein Beispiel: Wenn die Wahrscheinlichkeit (das Risiko), ohne Behandlung innerhalb eines Jahres zu sterben 90% (z.B. 9/10) beträgt und mit Behandlung 60% (z.B. 6/10), dann ergäbe die Zunahme des Risikos (ohne vs. mit Behandlung) ein $RR = 0,9/0,6 = 1,5$. Die entsprechende Odds Ratio wäre aber deutlich höher (mit den entsprechenden Odds von 9/1 und 6/4): $OR = (9 \cdot 4)/(6 \cdot 1) = 6$. Weil die Odds insbesondere bei hohen Wahrscheinlichkeiten sehr groß werden können, wird manchmal empfohlen, im Zweifelsfall eher relative Risiken zu benutzen, die meist auch leichter interpretiert werden können.

9.6.3 Mehr zu Effektgrößen in diesem Buch

Einige weitere Effektgrößen, insbesondere Maße der erklärten Varianz, werden wir in ► Teil III und ► Teil V dieses Buches beschreiben. Dort werden wir auch für jedes Verfahren zeigen, wie man Effektgrößen auf einfache Weise aus den Ergebnissen von sogenannten *Signifikanztests* berechnen kann. In ► Kapitel 27 widmen wir uns außerdem der Berechnung von *Konfidenzintervallen* für Effektgrößen. Wir werden im Zusammenhang mit entsprechenden Verfahren auch diskutieren, welchen Unterschied es für die Effektgrößenberechnung macht, ob ein Between-Subjects- oder ein Within-Subjects-Design (siehe ► Kapitel 5) verwendet wurde, ob man also zwei unabhängige Gruppen miteinander vergleicht oder *eine* Gruppe zu verschiedenen Zeitpunkten. Im Zusammenhang mit inferenzstatistischen Verfahren werden wir auch immer wieder auf Populations-Effektgrößen zur Berechnung der sogenannten *Teststärke* oder *Power* zurückkommen.

⁸ Auch das ist wieder ein Beispiel für ein Maß, dessen englische Bezeichnung auch im Deutschen verwendet wird (eine relativ selten verwendete deutsche Bezeichnung ist *Quotenverhältnis*).

ZUSAMMENFASSUNG

Effektgrößen gewinnen in der psychologischen Forschung immer mehr an Bedeutung, weil sie es erlauben, allgemeine und über Studien hinweg vergleichbare Aussagen über die Größe von Unterschieden, Zusammenhängen und anderen Arten von Resultaten zu machen. Grundsätzlich kann man Effektgrößen in zwei Arten von Maßen einteilen: Abstandsmaße und Zusammenhangsmaße. Abstandsmaße werden in der Regel in Standardabweichungseinheiten ausgedrückt, und Zusammenhangsmaße sind fast immer Variationen des Pearson-Korrelationskoeffizienten.

Beide Maße kann man aus Rohdaten berechnen, aber man kann die Maße auch ineinander überführen: Wenn man ein Abstandsmaß berechnet hat, kann man es leicht in ein Zusammenhangsmaß umrechnen (und umgekehrt). Die gebräuchlichsten Effektgrößen sind die Abstandsmaße d und g und die Korrelation (r). In manchen Fällen (z.B. zur Kommunikation von Effektgrößen für statistische Laien) kann es sinnvoll sein, r in Prozentwerten auszudrücken, als sogenannten *Binomial Effect Size Display* (BESD). In medizinischen Bereichen von Psychologie und Sozialwissenschaften, in denen oft nur dichotome Ergebnisse vorliegen, benutzt man häufig zwei weitere Effektgrößen: Relative Risiken (RR) und Odds Ratios (OR).

Weiterführende Literatur

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.). *The handbook of research synthesis*. New York: Russel Sage Foundation (231–244).

Rosnow, R. L. & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221–237.

Beides sind gut lesbare Übersichten, ohne besondere mathematischer Vorkenntnisse verständlich.

Tatsuoka, M. (1993). Effect size. In G. Keren & C. Lewis (Eds.). *A handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale, NJ: Erlbaum (461–479).

Dieses Kapitel verlangt etwas mehr mathematisches Vorwissen.

Teil 3: Inferenzstatistik

Kapitel 10	Grundlagen der Inferenzstatistik	325
Kapitel 11	Konfidenzintervalle	357
Kapitel 12	Signifikanztests	381
Kapitel 13	<i>t</i>-Tests	417
Kapitel 14	Der <i>F</i>-Test in der einfaktoriellen Varianzanalyse	437
Kapitel 15	Weitere <i>F</i>-Tests	469
Kapitel 16	Kontrastanalyse	511
Kapitel 17	Verfahren zur Analyse nominalskalierter Daten: Chi-Quadrat (χ^2)-Tests	539
Kapitel 18	Verfahren zur Analyse ordinalskalierter Daten	567
Kapitel 19	Resampling-Verfahren	583

Grundlagen der Inferenzstatistik

10

10.1	Wahrscheinlichkeiten, kurz gefasst	327
10.2	Von der Population über Stichproben zur Stichprobenverteilung	331
10.3	Stichprobenverteilung für Anteile	336
10.4	Lage- und Streuungsmaße von Stichprobenverteilungen	340
10.5	Der Einfluss der Stichprobengröße auf die Stichprobenverteilung	347
10.6	Rekapitulation und Ausblick	353

In den vorangegangenen Kapiteln dieses Buchs, die mit Datenanalyse zu tun hatten, haben wir uns durchweg mit Stichprobenergebnissen befasst. Psychologische und sozialwissenschaftliche Theorien beziehen sich jedoch äußerst selten auf Stichproben, sondern fast immer auf Populationen wie etwa Soziologiestudierende, Friseure, katholische Pfarrer, Kinder, Frauen, Männer oder einfach alle Menschen. In seltenen Fällen ist es zumindest im Prinzip möglich, alle Mitglieder einer Population zu untersuchen (z.B. alle Mitglieder des Bundestages), aber in der Regel scheitert die Untersuchung einer Population an großen praktischen Schwierigkeiten. Wie kann man trotzdem zu verwertbaren Ergebnissen kommen? Man versucht, aufgrund der Stichprobenergebnisse Schlussfolgerungen auf die Population zu ziehen. Genauer gesagt benutzt man Stichprobenstatistiken als Schätzungen für Populationsparameter. Beides sind Bezeichnungen für zusammengefasste Werte wie Anteile, Mittelwerte, Mittelwertsunterschiede oder Varianzen, die sich einmal auf Stichproben (Statistiken) und das andere Mal auf Populationen (Parameter) beziehen. Warum und wie man solche Schlüsse oder Inferenzen von der Stichprobe auf die Population ziehen kann, ist Gegenstand der *Inferenzstatistik*. Die zwei Arten von inferenzstatistischen Verfahren, mit denen wir uns in diesem Buch hauptsächlich beschäftigen werden, sind *Konfidenzintervalle* und *Signifikanztests*. Konfidenzintervalle dienen dazu, Genauigkeitsaussagen über Schätzungen von Populationsparametern zu treffen und Signifikanztests benutzt man, um Hypothesen über Populationsparameter zu prüfen. Dieses Kapitel behandelt die wahrscheinlichkeitstheoretischen Grundlagen der Inferenzstatistik, auf denen alle speziellen Verfahren, die in den nächsten Kapiteln beschrieben werden, beruhen.

Wir werden uns zunächst mit der Frage beschäftigen, warum Stichprobenstatistiken überhaupt als Schätzung für Populationsparameter brauchbar sind. Allerdings ist es mit der Schätzung von Populationsparametern nicht getan. Wie bekommt man zusätzlich Genauigkeitsaussagen über diese Populationsparameter? Wie kann man entscheiden, ob eine Hypothese über Populationsparameter (z.B. die Hypothese, dass sich die Mittelwerte zweier Populationen unterscheiden) zutrifft? Wir werden sehen, dass die theoretische Grundlage für solche Aussagen und Schlüsse – also für Konfidenzintervalle und Signifikanztests – sogenannte *Stichprobenverteilungen* sind.¹ Das sind theoretische Verteilungen von Stichprobenstatistiken, die Auskunft darüber geben, mit welcher Wahrscheinlichkeit man welche Stichprobenergebnisse erwarten kann, wenn man bestimmte Annahmen über die jeweilige Population trifft. Zur Frage, was das für Annahmen sind und wie sie sich für Konfidenzintervalle und Signifikanztests unterscheiden, erfahren Sie mehr in den zwei folgenden Kapiteln.

Nach einer kurzen Einführung in einige grundlegende Aspekte der Wahrscheinlichkeitstheorie, die wir später benötigen werden, erklären wir, wie man Stichprobenverteilungen für Anteile und für Mittelwerte erhält und was sie inhaltlich aussagen. Darauf zeigen wir, wie die Varianz und auch die Form der Stichprobenverteilung von der Stichprobengröße beeinflusst werden. Form und Varianz von Stichprobenverteilungen wiederum haben einen starken Einfluss auf alle Arten von inferenzstatistischen Aussagen. Zum Schluss wird der Inhalt dieses Kapitels noch einmal zusammenfassend betrachtet und in den Kontext der weiteren Kapitel über Inferenzstatistik gestellt. Für mathematisch interessierte Leser haben wir einige „Hintergrund-Kästen“ eingefügt, die beim ersten Lesen ohne große Einbußen im Verständnis übersprungen werden können.

¹ Stichprobenverteilungen werden in manchen Büchern als „Stichprobenkennwerteverteilungen“ bezeichnet und Stichprobenstatistiken als „Stichprobenkennwerte“.

Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwort- und DRM-Schutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: **info@pearson.de**

Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten oder ein Zugangscode zu einer eLearning Plattform bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.** ZugangsCodes können Sie darüberhinaus auf unserer Website käuflich erwerben.

Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

<https://www.pearson-studium.de>



Pearson