Forschungsmethoden und Statistik

für Psychologen und Sozialwissenschaftler

3., aktualisierte und erweiterte Auflage

Peter Sedlmeier
Frank Renkewitz

Pearson

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

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.

Z U S A M M E N F A S S U N G

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 medizinnahen 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).

Z U S A M M E N F A S S U N G

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 III

Inferenzstatistik

Grundlagen der Inferenzstatistik

10

ÜBERBLICK

» 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 Stichproben*statistiken* als Schätzungen für Populations*parameter*. 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. Danach 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

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

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. «

10.1 Wahrscheinlichkeiten, kurz gefasst

Es ist offensichtlich, dass Inferenzen oder Schlüsse über Populationsparameter nicht mit hundertprozentiger Sicherheit gezogen werden können, sondern immer potenziell fehlerbehaftet sind. Daran kann man nichts ändern, man kann aber versuchen, solche Schätzungen oder Urteile so genau und fundiert wie möglich zu machen. Hierzu muss man mit Wahrscheinlichkeiten operieren. Dieser Abschnitt gibt einen kurzen Überblick über einige Grundlagen der Wahrscheinlichkeitstheorie. Die Wahrscheinlichkeitstheorie ist ein eigener, sehr umfangreicher Wissenschaftsbereich mit einem teilweise hohen Formalitätsgrad. Wir beschränken uns aber auf den Wissensstoff, der nötig ist, um die gebräuchlichsten inferenzstatistischen Verfahren in Psychologie und Sozialwissenschaften zu verstehen. Die wahrscheinlichkeitstheoretischen Grundlagen der Inferenzstatistik werden dabei in einer möglichst intuitiven Weise (Sedlmeier, 1999; 2007) an einfachen Beispielen verdeutlicht. Trotzdem wird dieses Kapitel vielen Lesern wohl größere Schwierigkeiten bereiten als die vorangegangen. Der Aufwand lohnt sich jedoch: Die Prinzipien, die in diesem Kapitel dargestellt werden, gelten für alle Verfahren, die wir in späteren Kapiteln vorstellen werden. Wenn Sie also den Inhalt dieses Kapitels verstanden haben, sollte auch das Verstehen der in den folgenden Kapiteln beschriebenen Verfahren kein großes Problem darstellen.

10.1.1 Was ist Wahrscheinlichkeit?

Wir kennen den Begriff „Wahrscheinlichkeit“ aus dem Alltag: „Wahrscheinlich schreibe ich in der Klausur eine 2“, „Es ist sehr unwahrscheinlich, dass es heute regnen wird“, „Zu Ferienbeginn sind Staus auf der A4 sehr wahrscheinlich“. Woher kommen diese Wahrscheinlichkeiten und was bedeuten sie? Die Antworten auf die letzten beiden Fragen können sehr unterschiedlich ausfallen. Zum Wahrscheinlichkeitsbegriff gibt es eine umfangreiche philosophische Literatur, auf die wir nicht im Detail eingehen können. Aber eine häufig gemachte Unterscheidung ist die zwischen *subjektiver*, *klassischer* und *empirischer* oder *frequentistischer* Wahrscheinlichkeit.

Die Grundlage für eine subjektive Wahrscheinlichkeitsschätzung kann ein Gefühl oder eine Intuition sein, gemischt mit mehr oder weniger umfangreichem Hintergrundwissen: „Wie wahrscheinlich ist es, dass es menschenähnliche Lebewesen auf anderen Planeten gibt?“, wäre wohl eine Frage, auf die die meisten Menschen mit einer solch subjektiven Wahrscheinlichkeitsschätzung antworten würden. Wahrscheinlichkeiten variieren zwischen $p = 0$ und $p = 1$ (p steht für *probability*) oder zwischen 0% und 100%. Bei der Antwort auf die obige Frage würde wahrscheinlich der ganze Wertebereich ausgeschöpft, wenn man viele Personen befragte.

Deutlich mehr Übereinstimmung kann man erwarten, wenn man Wahrscheinlichkeiten nach dem sogenannten *klassischen Ansatz* bestimmt. Der lässt sich anwenden, wenn

man die Wahrscheinlichkeiten aus den Eigenschaften der Gegenstände ableiten kann. So hat ein Würfel beispielsweise sechs Seiten und es gibt keinen Grund, warum er (wenn es ein fairer Würfel ist) auf eine bestimmte Seite bevorzugt fallen sollte. Oder wenn man die Wahrscheinlichkeit dafür bestimmen möchte, dass eine Münze auf Kopf fällt, kann man aus der Überlegung, dass die Münze nur entweder auf Kopf oder Zahl fallen kann (nehmen wir an, dass die Münze nicht auf dem Rand stehen bleibt) und dass es wieder keinen Grund gibt, warum ein Ergebnis bevorzugt auftreten sollte, die entsprechende Wahrscheinlichkeit ableiten. Dies geht bei allen Arten von Glücksspielen (die auch die Grundlage für die Ableitung dieses Wahrscheinlichkeitsbegriffs durch Blaise Pascal waren). Generell kommt man zur entsprechenden Wahrscheinlichkeit eines Ereignisses A, indem man die Anzahl der „günstigen" Fälle durch die Anzahl der möglichen Fälle teilt, die sich wiederum zusammensetzen aus der Anzahl der günstigen und der nicht-günstigen Fälle. Die Wahrscheinlichkeit von A, abgekürzt als $p(A)$ ist also:

$$p(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

Wenn man beispielsweise eine „6" würfeln möchte, gibt es einen günstigen Fall (die „6" zu würfeln) und fünf ungünstige Fälle (1, 2, 3, 4 oder 5 als Würfelergebnis zu erhalten). Die Wahrscheinlichkeit, eine „6" zu würfeln, ist also $p = 1/(1 + 5) = 1/6$ und die Wahrscheinlichkeit, dass die Münze auf Kopf fällt, ist $p = 1/(1 + 1) = 1/2$.

Der Anwendungsbereich für diese klassische Vorgehensweise ist offensichtlich sehr eingeschränkt, weil die Wahrscheinlichkeit der Einzelereignisse eindeutig aus den Eigenschaften der betreffenden Gegenstände bestimmbar sein muss. Subjektive Wahrscheinlichkeiten können demgegenüber zwar prinzipiell immer geschätzt werden, aber oft wird es große Uneinigkeit über die genauen Werte geben. Deswegen benutzt man in der Praxis meist empirische oder *frequentistische* Wahrscheinlichkeiten. Man schätzt sie aufgrund von relativen Häufigkeiten. Die Grundlage für die Häufigkeitswerte sind jedoch nicht theoretische Überlegungen, sondern Beobachtungen. Die relative Häufigkeit erhält man, indem man die Anzahl der „passenden" Beobachtungen $f(A)$ (f steht für *frequency*) durch die Anzahl aller relevanten Beobachtungen, der passenden und der nicht-passenden $f(\neg A)$ (¬ steht für „nicht"), teilt:

$$p(A) = \frac{f(A)}{f(A) + f(\neg A)}$$

Wenn wir z.B. die Beobachtung gemacht hätten, dass in einer Zufallsstichprobe von 20 Frauen sechs an Depressionen leiden und 14 nicht, dann wäre unsere Schätzung für die Wahrscheinlichkeit, dass eine zufällig ausgewählte Frau aus der entsprechenden Population an Depressionen leidet, $p = 6/(6 + 14) = 0{,}3$ oder 30%. Die Wahrscheinlichkeiten, mit denen wir uns in diesem Buch befassen, werden wir weitgehend auf diese Art ableiten. Besonders wichtig ist bei frequentistischen Wahrscheinlichkeiten die Rolle des Zufalls. Während der Zufall bei klassischen Wahrscheinlichkeiten gewissermaßen „automatisch" mitspielt, muss man bei frequentistischen Wahrscheinlichkeiten darauf achten, dass die Häufigkeiten, die man zur Wahrscheinlichkeitsschätzung benutzt, durch eine Zufallsstichprobe entstanden sind oder zumindest so interpretiert werden können.

10.1.2 Wahrscheinlichkeit von Konjunktionen und bedingte Wahrscheinlichkeiten

Wahrscheinlichkeitsschätzungen beziehen sich häufig auf zusammengesetzte Ereignisse oder *Konjunktionen*, wie z.B. „Die Sonne scheint *und* es ist warm“ oder „Das Essen schmeckt, *aber* es ist teuer“: *und* und *aber* zeigen hier an, dass beide Ereignisse jeweils zusammen zutreffen müssen. Wir hatten es in *Kapitel 1* (*Abschnitt 1.1.4*) schon einmal mit einer Konjunktion von Ereignissen zu tun (die Aussage b in folgendem Problem):

> *Ein Schüler hat im Abschlusszeugnis in Mathematik die Note Vier erzielt. Welche der folgenden Aussagen trifft für ihn mit größerer Wahrscheinlichkeit zu?*
>
> *a) Er hatte eine Sechs in Mathe im Halbjahreszeugnis.*
>
> *b) Er hatte eine Sechs in Mathe im Halbjahreszeugnis, hat aber im zweiten Halbjahr Nachhilfe in Mathe erhalten.*

▶ Abbildung 10.1 übersetzt das Problem in eine grafische Darstellung. Die 30 Kästchen repräsentieren 30 Schüler, die im Abschlusszeugnis eine Vier hatten. Von diesen haben zwölf Nachhilfe bekommen (dunkle Kästchen) und acht hatten eine Sechs im Halbjahreszeugnis (Kästchen mit einem Kreuz). Bei fünf Schülern traf beides zu (dunkle Kästchen mit Kreuz). Die Abbildung macht eine Eigenheit der Wahrscheinlichkeit von Konjunktionen deutlich: Ihre Wahrscheinlichkeit kann nie größer sein als die Wahrscheinlichkeit jedes der Einzelereignisse. In diesem Beispiel ist die Wahrscheinlichkeit dafür, dass ein (zufällig ausgewählter) Schüler eine Sechs im Halbjahreszeugnis hatte, 8/30 und die, dass ein Schüler Nachhilfe bekam, ist 12/30. Die Wahrscheinlichkeit der Konjunktion ist aber nur 5/30. Die Anzahl der Schüler in ▶ Abbildung 10.1, für die eine bestimmte Information zutrifft, wurde willkürlich gewählt. Die Schlussfolgerung stimmt aber immer: Die Wahrscheinlichkeit einer Konjunktion kann nie größer sein als die Wahrscheinlichkeit jeder der Einzelwahrscheinlichkeiten.

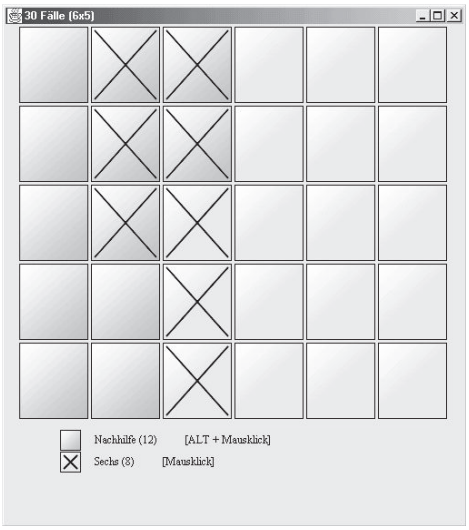

Abbildung 10.1: Illustration des Sachverhalts, dass die Wahrscheinlichkeit einer Konjunktion von Ereignissen nie größer sein kann als die Wahrscheinlichkeit jedes der Einzelereignisse. Die 30 Kästchen repräsentieren 30 Schüler mit einer Vier in Mathe im Abschlusszeugnis (erstellt mit dem Trainingsprogramm aus Sedlmeier & Köhlers, 2001).

Die Wahrscheinlichkeit von Konjunktionen ist für sich gesehen schon interessant, aber man braucht sie auch, um *bedingte Wahrscheinlichkeiten* zu bestimmen. In einfachen Fällen erkennt man die Bedingung an dem Wort *wenn*, das die Bedingung spezifiziert: Wie groß ist die Wahrscheinlichkeit, dass ein Schüler in ▶ Abbildung 10.1 Nachhilfe bekommen hat, *wenn* er eine Sechs im Halbjahreszeugnis hatte? Das ist nichts anderes als der Anteil der Schüler, auf die beides zutrifft (Nachhilfe *und* Sechs – eine Konjunktion), an den Schülern, die eine Sechs hatten (die Bedingung): 5 von 8 oder 5/8. Generell kann man eine bedingte Wahrscheinlichkeit für das Ereignis A unter der Bedingung B mit folgender Formel berechnen (rechts von dem senkrechten Strich „|“ steht die Bedingung und das Zeichen „∧“ steht für „und“):

$$p(A|B) = \frac{p(A \wedge B)}{p(B)}$$

Eine bedingte Wahrscheinlichkeit für A gegeben B erhält man also, wenn man die Wahrscheinlichkeit der Konjunktion von A und B durch die Wahrscheinlichkeit der Bedingung B teilt. Man kann zur Berechnung direkt die zur Verfügung stehenden Häufigkeiten benutzen (siehe ▶ Abbildung 10.1). So ist beispielsweise die Wahrscheinlichkeit, dass ein Schüler eine Sechs schreibt, wenn er Nachhilfe bekam:

$$p(Sechs|Nachhilfe) = \frac{f(Sechs \wedge Nachhilfe)}{f(Nachhilfe)} = \frac{5}{12}$$

Man kann auch zuerst die Wahrscheinlichkeiten für Zähler und Nenner schätzen und dann damit weiter rechnen (insbesondere, wenn diese Wahrscheinlichkeiten schon zur Verfügung stehen, wird man so verfahren):

$$p(Sechs \wedge Nachhilfe) = \frac{f(Sechs \wedge Nachhilfe)}{f(alle\ Schüler)} = \frac{5}{30}$$

$$p(Nachhilfe) = \frac{f(Nachhilfe)}{f(alle\ Schüler)} = \frac{12}{30}$$

$$p(Sechs|Nachhilfe) = \frac{p(Sechs \wedge Nachhilfe)}{p(Nachhilfe)} = \frac{\frac{5}{30}}{\frac{12}{30}} = \frac{5}{12}$$

Bedingte Wahrscheinlichkeiten kommen im Alltag häufig vor, sind aber nicht immer leicht zu erkennen, wie bei dem folgenden Beispiel, mit dem wir es schon einmal zu tun hatten:

Welches der folgenden Ereignisse ist wahrscheinlicher?

a) Eine Hausfrau hat promoviert.

b) Eine promovierte Frau ist Hausfrau.

Ein Problem ist hier, erst einmal zu erkennen, dass es sich um bedingte Wahrscheinlichkeiten handelt. Übersetzt in eine formale Notation wären die Aussagen *a* und *b*:

a) p(promoviert | Hausfrau)

b) p(Hausfrau | promoviert)

Copyright