COMPANION WEBSITE
CWS
www.pearson-studium.de

ps
psychologie

Peter Sedlmeier
Frank Renkewitz

Forschungsmethoden und Statistik in der Psychologie

PEARSON
Studium

Korrelation

7

ÜBERBLICK

Ganz offensichtlich besteht ein Zusammenhang zwischen der Körpergröße und dem Körpergewicht von Personen: Größere Personen sind im Allgemeinen schwerer als kleinere. Ebenso offensichtlich ist dieser Zusammenhang nicht perfekt: Wenn wir wissen, dass Herr Kunze 1,82 m groß ist, können wir sein Gewicht nicht präzise angeben und wir werden auch kaum annehmen, dass die größte Person in einem Raum zwingend zugleich die schwerste ist. Der generelle Trend ist dennoch eindeutig: Niedrige Werte auf der Variablen Körpergröße gehen mit eher niedrigen Werten auf der Variablen Körpergewicht einher, während höhere Werte bei der Körpergröße zumeist auch mit höheren Werten beim Körpergewicht verbunden sind. Die Beziehung zwischen den Merkmalen Körpergröße und Körpergewicht ist somit ein Beispiel für eine Korrelation – denn der Begriff Korrelation bezeichnet nichts anderes als einen Zusammenhang zwischen Variablen.

Die Suche nach Korrelationen und die Analyse solcher Korrelationen ist ein Hauptanliegen aller empirischen Wissenschaften. In der psychologischen Forschung können fast alle Fragestellungen auch als Fragen nach dem Zusammenhang zwischen Variablen aufgefasst werden. Haben intelligentere Eltern intelligentere Kinder? Gibt es einen Zusammenhang zwischen der Intelligenz und dem Erfolg im Berufsleben? Steigt die Leistung bei einer Aufgabe mit der Motivation für diese Aufgabe? Sinkt der Lernerfolg – wie von vielen Lehrern vermutet – mit zunehmender Klassengröße? Sind zufriedenere Arbeitnehmer seltener krank als weniger zufriedene Arbeitnehmer? Verbessert sich das Befinden Depressiver mit der Anzahl ihrer Sozialkontakte oder mit der Teilnahme an einer Verhaltenstherapie? Sind extravertierte Menschen beliebter als introvertiertere? Gibt es einen Zusammenhang zwischen der Attraktivität von Personen und ihrer Treue in Beziehungen?

Selbstverständlich ist es ein wesentliches Ziel der wissenschaftlichen Psychologie, solche Fragen zu beantworten und somit Zusammenhänge zwischen Variablen aufzudecken. Die Analyse der Beziehung zwischen verschiedenen Merkmalen erschöpft sich allerdings nicht darin zu klären, ob überhaupt ein Zusammenhang besteht. Relevant ist darüber hinaus die Richtung, Stärke und Form des Zusammenhangs. Was mit diesen Aspekten von Korrelationen genau gemeint ist und wie Korrelationen mit Hilfe von statistischen Verfahren umfassend analysiert werden können, werden wir im Verlauf dieses Kapitels erläutern. Im Gegensatz zum vorangegangenen Kapitel, in dem Möglichkeiten zur Beschreibung *einer* Variablen erörtert wurden, wenden wir uns damit Methoden zu, mit denen Beziehungen zwischen *zwei* Variablen beschrieben werden können. Wir werden uns zunächst ansehen, wie Korrelationen mit Hilfe von so genannten Streudiagrammen grafisch dargestellt werden können. Diese Streudiagramme werden wir dann dazu nutzen, verschiedene Formen von Zusammenhängen zu veranschaulichen und die Bedeutung der Stärke und Richtung eines Zusammenhangs zu klären. Anschließend werden wir den *Produkt-Moment-Korrelationskoeffizienten* kennen lernen – ein Maß für die Stärke und Richtung des Zusammenhangs zwischen zwei intervallskalierten Variablen. Wir werden zudem erörtern, wie Korrelationen interpretiert werden können und dabei besonders auf mögliche Verfälschungen des Korrelationskoeffizienten und das Verhältnis von Korrelation und Kausalität eingehen. Wir werden uns dann der Partialkorrelation zuwenden – einem Verfahren, mit dem untersucht werden kann, wie der Zusammenhang zwischen zwei Variablen durch eine dritte Variable beeinflusst wird. Das Kapitel endet mit einer Einführung in den Phi-Koeffizienten und Spearman's Rho. Dies sind Maße, die eingesetzt werden können, wenn der Zusammenhang zwischen Variablen, die nicht mindestens Intervallskalenniveau aufweisen, bestimmt werden soll.

7.1 Die grafische Darstellung von Korrelationen: Streudiagramme

Medienberichte über Gewalt unter Kindern und Jugendlichen werfen häufig die Frage auf, ob es einen Zusammenhang zwischen der Aggressivität von Kindern und ihrem Konsum gewaltdarstellender Fernsehprogramme gibt. Wie könnten wir diese Frage untersuchen? Zunächst müssten wir in einer Gruppe von Kindern den Wert beider interessierenden Variablen bestimmen. Es wäre also für jedes Kind zu ermitteln, in welchem Umfang es gewaltdarstellende Sendungen sieht *und* wie aggressiv es sich verhält. In typischen Studien zu dieser Fragestellung werden die teilnehmenden Kinder z.B. danach gefragt, welche Sendungen sie sich wie häufig ansehen. Diese Sendungen werden dann daraufhin bewertet, wie gewalttätig sie sind. Auf diese Weise kann ein Index für den Konsum Gewalt darstellender Sendungen gebildet werden, der etwa von 0 (kein Konsum) bis 10 (starker Konsum) reichen könnte. Die Aggressivität der Kinder könnte von Lehrern und Klassenkameraden auf Rating-Skalen mit einem Wertebereich von 0 (nicht aggressiv) bis 7 (äußerst aggressiv) beurteilt werden.

Nehmen wir an, wir hätten eine solche Studie mit 20 Kindern durchgeführt. Der erste Schritt bei der Auswertung der Daten sollte dann darin bestehen, ein Streudiagramm zu erstellen. Mit diesem Streudiagramm können wir auf einen Blick einen guten ersten Eindruck davon gewinnen, ob zwischen den untersuchten Variablen ein Zusammenhang besteht und wie stark dieser Zusammenhang ausgeprägt ist. Die (fiktiven) Ergebnisse unserer Studie sind im Streudiagramm in ▶Abbildung 7.1 dargestellt. Wie man sieht, besteht ein Streudiagramm aus einem einfachen Koordinatensystem. Auf der x-Achse (oder Abszissenachse) wird die eine Variable abgetragen, die andere Variable findet sich auf der y-Achse (oder Ordinatenachse). Bei uns sei der „Konsum Gewalt darstellender Fernsehsendungen“ die Variable X und die Aggressivität die Variable Y. Jeder Punkt im Streudiagramm gibt nun die Werte eines Kindes auf beiden Variablen an. In der Abbildung sind die Daten eines Kindes mit den Werten $x = 3$ und $y = 4$ besonders hervorgehoben. Für dieses Kind wurde also hinsichtlich des Konsums gewalttätiger Sendungen ein Wert von 3 ermittelt, während seine Aggressivität mit 4 beurteilt wurde.

Das generelle Muster in unseren Daten ist recht eindeutig. Die „Punktwolke“ verläuft von links unten nach rechts oben. Demnach sind Kinder, die eher wenig Gewalt darstellende Sendungen sehen, tendenziell auch weniger aggressiv als Kinder, die häufig Gewaltdarstellungen konsumieren. Augenscheinlich ist also ein Zusammenhang vorhanden. Das Streudiagramm zeigt allerdings auch, dass dieser Zusammenhang nicht perfekt ist. Es finden sich zahlreiche Kinder mit gleichen oder ähnlichen Werten beim Fernsehkonsum, die sich aber hinsichtlich ihrer Aggressivität deutlich unterscheiden. Unsere fiktiven Daten entsprechen damit in etwa dem typischen Befund in vergleichbaren Studien (z.B. Eron, 1982; Huesmann, 1982): Es gibt bei Kindern und Jugendlichen tatsächlich einen recht ausgeprägten Zusammenhang zwischen der konsumierten Menge an Gewalt im Fernsehen und der Aggressivität.

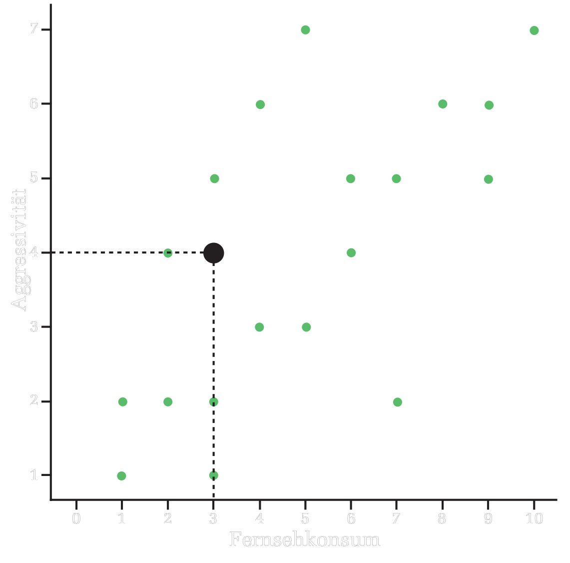

Abbildung 7.1: Ein Streudiagramm, das den Zusammenhang zwischen dem „Konsum Gewalt darstellender Fernsehsendungen" und der „Aggressivität" bei Kindern illustriert (fiktive Daten).

In der Regel ist es gleichgültig, welche der beteiligten Variablen in einem Streudiagramm auf der x-Achse und welche auf der y-Achse eingezeichnet wird. Anders verhält es sich allerdings in Studien, in denen eine Variable zur Vorhersage der anderen verwendet werden soll oder in denen angenommen wird, dass eine Variable die andere verursacht. In diesen Fällen wird die vorhersagende oder verursachende Variable in aller Regel mit X bezeichnet und entsprechend eingezeichnet. Würden wir also den Zusammenhang zwischen Abiturnote und Studienerfolg untersuchen, so würde die Abiturnote auf der x-Achse abgetragen, da es nahe liegt, diese Variable auch zur Vorhersage des Studienerfolgs zu nutzen. In unserer Beispielstudie könnte die Entscheidung, den Konsum Gewalt darstellender Fernsehsendungen auf der x-Achse einzuzeichnen, durch die Annahme begründet sein, dass diese Variable eine Ursache für erhöhte Aggressivität ist. Es ist allerdings äußerst wichtig zu beachten, dass auch ein starker Zusammenhang zwischen den Variablen *nicht* belegt, dass der Konsum von Gewaltdarstellungen tatsächlich Aggressivität verursacht. Wir werden auf das Verhältnis von Kausalität und Korrelation im Abschnitt 7.5 zurückkommen.

Im Kasten „Korrelationen und Sonnenblumen" wird eine Variante von Streudiagrammen vorgestellt.

HINTERGRUND

Korrelationen und Sonnenblumen Natürlich hätten wir in der oben geschilderten Untersuchung auch auf zwei Kinder mit identischen Messwertpaaren treffen können – also auf zwei Kinder, die sowohl hinsichtlich des Fernsehkonsums als auch hinsichtlich der Aggressivität dieselben Werte aufweisen. Generell treten identische Messwertpaare insbesondere dann auf, wenn wir große Stichproben untersuchen und wenn wir Variablen messen, die nur wenige verschiedene Werte annehmen können (dies ist z.B. bei Messungen mit Rating-Skalen der Fall). Im Streudiagramm werden identische Messwertpaare als übereinander liegende Punkte eingezeichnet. Man kann im Streudiagramm daher nicht feststellen, wie häufig ein bestimmtes Messwertpaar beobachtet wurde. Sofern in einer Studie viele Personen identische Messwerte haben, kann dies dazu führen, dass nur schlecht zu erkennen ist, ob zwischen den untersuchten Variablen ein Zusammenhang besteht. Abhilfe schafft in einem solchen Fall ein „Sonnenblumendiagramm“. Auch im Sonnenblumendiagramm wird für Messwertpaare, die nur einmal vorkommen, ein Punkt eingezeichnet. Haben jedoch mehrere Personen die gleichen Messwerte, so wird für jede dieser Personen ein „Blütenblatt“ in das Diagramm eingefügt. (Diese Blütenblätter werden in der Regel jedoch nur als Striche dargestellt.)

Das Sonnenblumendiagramm in ▸Abbildung 7.2 zeigt die hypothetischen Ergebnisse einer Studie zum Zusammenhang zwischen dem Interesse an Jazz und dem Interesse an klassischer Musik. Beide Variablen wurden mit Hilfe von vierstufigen Rating-Skalen gemessen (1 = „kein Interesse“, 4 = „sehr starkes Interesse“).

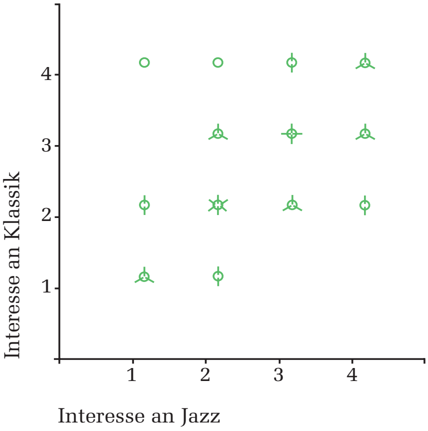

Abbildung 7.2: Ein Sonnenblumendiagramm für den Zusammenhang zwischen dem Interesse an Jazz und dem Interesse an Klassik.

Wie ist dieses Sonnenblumendiagramm nun zu lesen? Beispielsweise bezifferte eine Person ihr Interesse an Jazz mit 1, kreuzte aber beim Interesse an klassischer Musik den Wert 4 an. Zwei Teilnehmer wählten beim Interesse an Jazz den Wert 4 und beim Interesse an klassischer Musik den Wert 2. Sechs Teilnehmer gaben sowohl beim Interesse an Jazz als auch beim Interesse an klassischer Musik den Wert 2 an. Vier Teilnehmer wählten bei beiden Variablen den Wert 3 – und so weiter. Insgesamt wird im Sonnenblumendiagramm ein Zusammenhang deutlich, der im Streudiagramm kaum zu erkennen gewesen wäre: Zwar tritt bei jedem Interesse an Jazz nahezu jeder beliebige Wert für das Interesse an klassischer Musik auf. Bei größerem Interesse an Jazz häufen sich jedoch auch höhere Werte für das Interesse an Klassik, während bei geringem Jazz-Interesse zumeist auch das Interesse an Klassik nur schwach ausgeprägt ist.

7.2 Korrelationsmuster

7.2.1 Lineare und kurvilineare Zusammenhänge

Die Korrelation zwischen dem Konsum gewaltdarstellender Fernsehsendungen und der Aggressivität ist ein Beispiel für einen linearen Zusammenhang. Ein weiteres Beispiel für einen linearen Zusammenhang zeigt die ▶Abbildung 7.3a, die die Korrelation zwischen der Menge gelesener Bücher und Zeitschriften und dem Umfang des Allgemeinwissens bei Studierenden des ersten Semesters veranschaulicht (auch diese Daten sind erfunden, lehnen sich aber lose an einen realen Befund an, Stanovich & Cunningham, 1993). Das wesentliche Kennzeichen linearer Zusammenhänge besteht darin, dass das Muster der Datenpunkte im Streudiagramm annäherungsweise durch eine Gerade beschrieben werden kann. Eine solche Gerade haben wir in Abbildung 7.3a eingezeichnet. (Die genaue Bezeichnung lautet Regressionsgerade. Wir werden in Kapitel 8 im Detail klären, wie man Regressionsgeraden konstruiert.) Offensichtlich liegen nicht alle Punkte exakt auf der Gerade, einige weichen sogar stark davon ab. Dennoch gibt die Gerade den generellen Trend in den Daten recht gut wieder. Das Gleiche gilt für die Daten in Abbildung 7.1 zum Zusammenhang von Fernsehkonsum und Aggressivität. Auch hier könnten Sie eine Gerade einzeichnen, und diese würde den allgemeinen Verlauf der Datenpunkte gut beschreiben.

Selbstverständlich muss die Form des Zusammenhangs zwischen zwei Variablen nicht grundsätzlich linear sein. Es gibt auch in der Psychologie zahlreiche Beispiele für Beziehungen zwischen Variablen, die in ihrem Verlauf keiner Geraden entsprechen. ▶Abbildung 7.3b zeigt solch einen kurvilinearen Zusammenhang. Sie veranschaulicht die Beziehung zwischen dem Grad der physiologischen Aktivierung und der Leistung bei einer Aufgabe. Die Art der Aufgabe spielt dabei nahezu keine Rolle. Die hier skizzierte Beziehung ist bei sehr unterschiedlichen Aufgaben gültig, also etwa beim Lösen einer Statistikklausur ebenso wie beim Vorspielen mit einem Musikinstrument. Von einem besonders entspannten, fast schläfrigen Zustand bis zu einem mittleren Aktivierungsniveau steigt die Leistung. Jenseits dieses Niveaus nimmt die Leistung aber wieder ab – man ist schlicht zu aufgeregt, um ein optimales Ergebnis zu erzielen. Die Beziehung zwischen der physiologischen Aktivierung und der Leistung kann also offensichtlich deutlich besser durch eine Kurve beschrieben werden als durch eine Gerade. Nach der Form dieser Kurve wird diese Beziehung auch als umgekehrt U-förmiger Zusammenhang bezeichnet. Das Streudiagramm in ▶Abbildung 7.3c illustriert einen anderen kurvilinearen Zusammenhang. Hier geht es um die Beziehung zwischen der Größe eines Geldvermögens und dem subjektiven Nutzen dieses Vermögens. Wenn Sie keinerlei Geld besitzen, wird Ihnen eine Einnahme von 100 Euro wahrscheinlich als großes Glück erscheinen. Verfügen sie bereits über 10.000 Euro, so wären zusätzliche 100 Euro wohl immer noch ein netter Gewinn, aber schon ein weit geringerer Anlass zu übermäßiger Freude. Sofern ihr Kontostand sich auf 2 Millionen Euro beläuft, werden Sie es kaum noch bemerken, ob weitere 100 Euro eingehen oder nicht. Der Zuwachs im subjektiven Nutzen ihres Vermögens, der mit 100 Euro verbunden ist, wird also immer geringer je mehr Geld Sie schon haben.

Daher ergibt sich auch zwischen dem Geldvermögen und dem Nutzen des Vermögens ein Zusammenhang, der keiner Geraden entspricht, sondern einer Kurve.

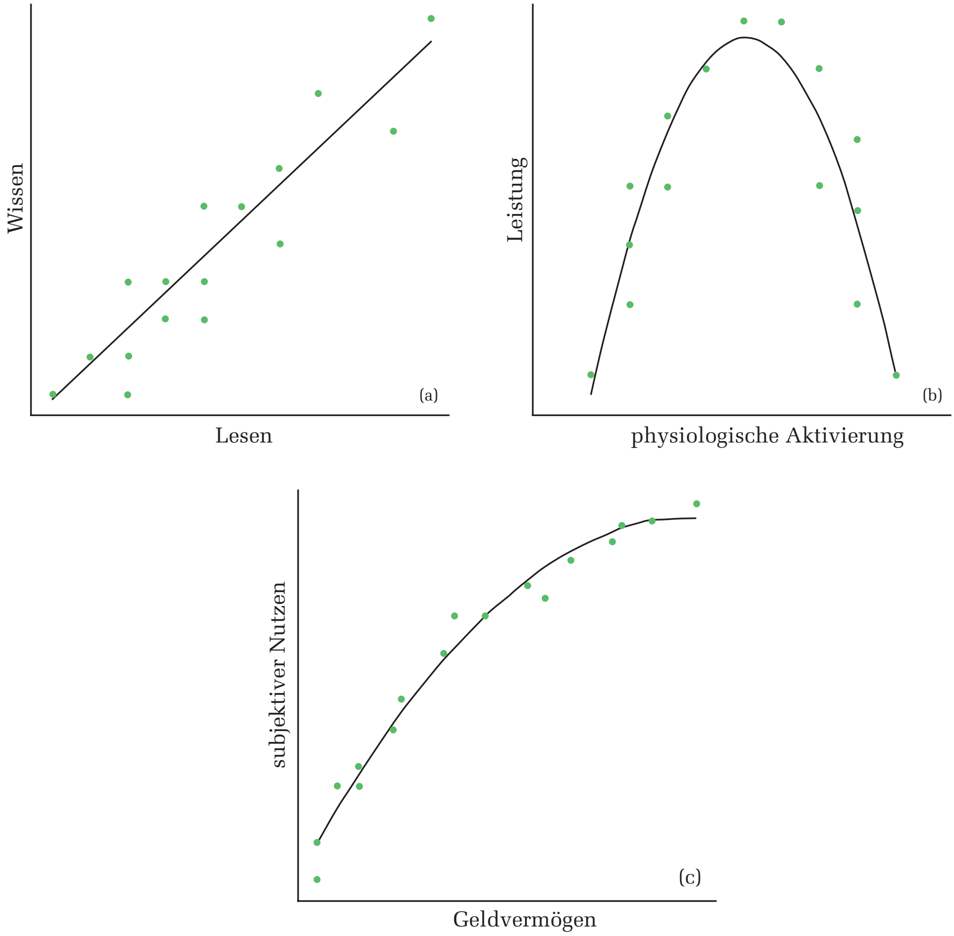

Abbildung 7.3: Unterschiedliche Formen von Zusammenhängen im Streudiagramm. Das Diagramm (a) zeigt einen linearen Zusammenhang, die Diagramme (b) und (c) illustrieren kurvilineare Zusammenhänge.

7.2.2 Richtung und Stärke von Zusammenhängen

Bei linearen Zusammenhängen können verschiedene Richtungen unterschieden werden. Die Richtung eines linearen Zusammenhangs ist entweder positiv oder negativ. Bei unseren bisherigen Beispielen handelt es sich um positive Korrelationen. Hier gehen jeweils hohe Werte auf der einen Variablen mit hohen Werten auf der anderen Variablen einher. Beispielsweise verfügen Studierende des ersten Semester, die viel gelesen haben, zumeist auch über ein großes Allgemeinwissen. Dagegen sind bei negativen Korrelationen hohe Werte auf einer Variablen mit niedrigen Werten auf der anderen Variablen verbunden. Ein Beispiel für eine negative Korrelation ist die Beziehung zwischen der Arbeitszufriedenheit von Arbeitnehmern und der Anzahl ihrer Fehltage: Beschäftigte mit geringerer Zufriedenheit erscheinen häufiger nicht an ihrem Arbeitsplatz (▶Abbildung 7.4). Die Bezeichnungen positiv und negativ für die Richtung linearer Zusammenhänge gehen auf die Steigung der Gerade zurück, mit der diese Zusammenhänge beschrieben werden können. Verläuft diese Gerade im Streudiagramm von links unten nach rechts oben, so hat sie eine positive Steigung. Fällt Sie dagegen von links nach rechts ab, so ist ihre Steigung negativ.

Abbildung 7.4: Negativer Zusammenhang zwischen der Arbeitszufriedenheit und der Anzahl der Fehltage.

Neben der Form und Richtung einer Korrelation lässt sich mit Hilfe eines Streudiagramms auch ihre Stärke beurteilen. Generell ist ein Zusammenhang zwischen zwei Variablen umso stärker ausgeprägt, je genauer wir bei Kenntnis des Werts einer Variablen angeben können, welchen Wert die andere Variable annehmen wird. Unabhängig von der Form des Zusammenhangs zeigt sich dies im Streudiagramm dadurch, dass die Datenpunkte bei starken Korrelationen näher an der gedachten Gerade oder Kurve liegen, die die Beziehung zwischen den Variablen am besten repräsentiert. Im Fall eines perfekten Zusammenhangs befinden sich alle Datenpunkte exakt auf dieser Gerade (siehe Abbildung 7.5a) oder Kurve. Perfekte Korrelationen werden auch als deterministische Zusammenhänge bezeichnet. Solche Zusammenhänge bestehen etwa zwischen der Anzahl der Biere, die Sie in Ihrem Lieblingsclub am Samstag trinken, und Ihrem Rechnungsbetrag (vorausgesetzt, Sie bestellen ausschließlich Bier) oder zwischen der Entfernung eines Sterns von der Erde und der Zeit, die das Licht von diesem Stern benötigt, um die Erde zu erreichen. Wir können für jede beliebige Anzahl von Bieren exakt und fehlerfrei angeben, was Sie bezahlen müssen. Ebenso lässt sich für jede Entfernung eines Sterns eindeutig bestimmen, wie lange das Licht von dort bis zur Erde unterwegs war. In der sozialwissenschaftlichen Forschung kommen solche deterministischen Zusammenhänge jedoch praktisch niemals vor. Hier finden wir stattdessen stochastische (oder probabilistische) Zusammenhänge. Bei stochastischen Korrelationen ist einem bestimmten Wert auf einer Variable X kein eindeutiger Wert auf der Variable Y zugeordnet. Vielmehr werden sich auch Personen, die auf der Variable X den gleichen Wert aufweisen, hinsichtlich ihrer Merkmalsausprägung in der Variable Y mehr oder weniger stark unterscheiden. Wir können etwa bei zwei Kindern, die im gleichen Umfang gewaltdarstellende Fernsehsendungen sehen, nicht damit rechnen, dass sie auch exakt gleich aggressiv sind. Studierende, die bisher gleich viel gelesen haben, werden sich hinsichtlich des Umfangs ihres Allgemeinwissens häufig dennoch zumindest geringfügig unterscheiden. Entsprechend weichen die Datenpunkte in Streudiagrammen, die solche linearen stochastischen Zusammenhänge veranschaulichen, mehr oder weniger stark von einer Gerade ab. Je weiter die Datenpunkte von der Gerade entfernt sind, desto schwächer ist der Zusammenhang zwischen den untersuchten Variablen. Besteht zwischen zwei Variablen schließlich überhaupt kein Zusammenhang, so findet sich im Streudiagramm auch kein Trend, der durch eine Gerade oder irgendeine Kurve repräsentiert werden könnte. Untersuchten wir beispielsweise die Korrelation zwischen dem Körpergewicht und der Musikalität von Personen, so würden die resultierenden Daten wahrscheinlich denen in ▶Abbildung 7.5b ähneln. Die Punktwolke ist etwa kreisförmig und hohe und niedrige Werte beim Körpergewicht gehen gleichermaßen mit hohen und niedrigen Werten bei der Musikalität einher. Demgemäß ermöglicht uns die Kenntnis des Körpergewichts einer Person auch keinerlei Aussage über ihre Musikalität; natürlich können wir ebenso wenig aus der Musikalität einer Person auf ihr Körpergewicht schließen.

Die weiteren Streudiagramme in ▶Abbildung 7.5 veranschaulichen noch einmal unterschiedlich starke positive und negative lineare Korrelationen. Es ist sehr nützlich eine gewisse Routine darin zu entwickeln, solche Streudiagramme zu „lesen“ und so z.B. sehr schnell eine Aussage darüber treffen zu können, ob zwischen zwei Variablen nur ein geringer oder ein außerordentlich hoher Zusammenhang besteht. Zu jedem Streudiagramm haben wir auch ein Maß für die Stärke und Richtung des jeweiligen linearen Zusammenhangs angegeben. Bei diesem Maß handelt es sich um den Produkt-Moment-Korrelationskoeffizienten, dessen Berechnung und Interpretation wir in Kürze genauer besprechen werden. Damit die Zahlen in Abbildung 7.5 nicht bedeutungslos bleiben, wollen wir allerdings schon jetzt den Wertebereich des Produkt-Moment-Korrelationskoeffizienten beschreiben: Der Korrelationskoeffizient kann Werte zwischen -1 und $+1$ annehmen. Das Vorzeichen zeigt dabei die Richtung des Zusammenhangs an. Bei negativen Zusammenhängen liegt der Korrelationskoeffizient also zwischen 0 und -1, bei positiven Zusammenhängen entsprechend zwischen 0 und $+1$. Die Stärke eines Zusammenhangs wird dagegen durch den Betrag des Korrelationskoeffizienten angegeben. Ein Korrelationskoeffizient von -1 bezeichnet also ebenso wie ein Korrelationskoeffizient von $+1$ einen perfekten Zusammenhang. Den Wert 0 nimmt der Korrelationskoeffizient an, wenn zwischen zwei Variablen kein linearer Zusammenhang besteht.

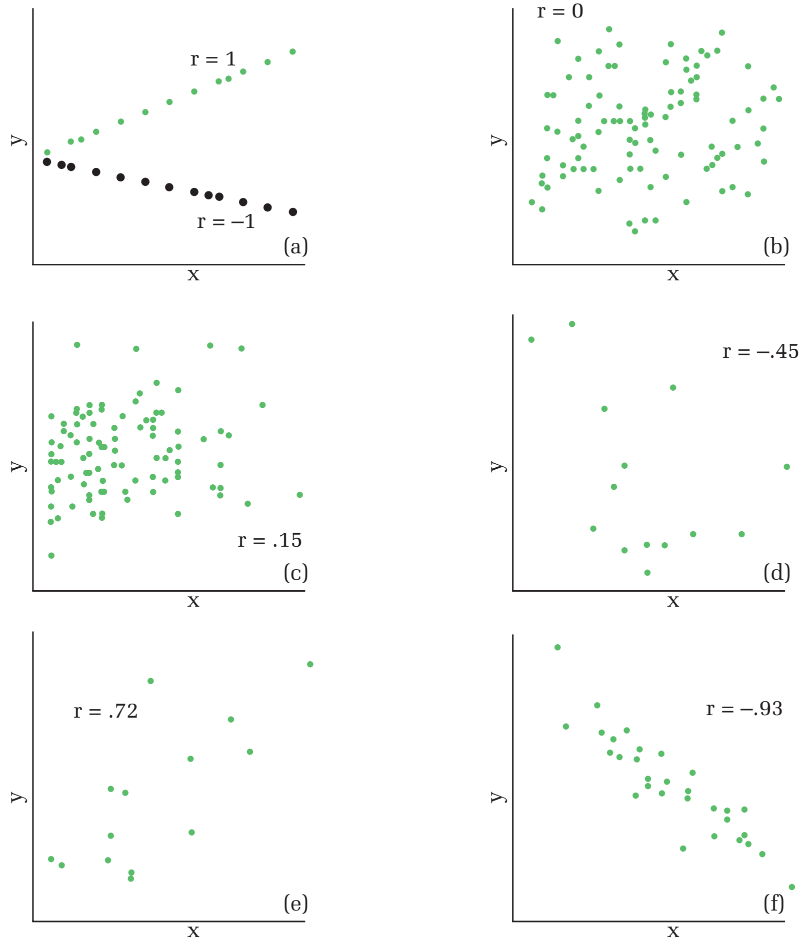

Abbildung 7.5: Streudiagramme und Produkt-Moment-Korrelationskoeffizienten bei Zusammenhängen unterschiedlicher Stärke und Richtung.

7.2.3 Die Bedeutung des Korrelationsmusters für die weitere Analyse

Das übliche Vorgehen bei der Analyse der Stärke des Zusammenhangs zwischen zwei Variablen besteht darin, den Produkt-Moment-Korrelationskoeffizienten zu bestimmen. Allerdings sollte vor der Berechnung dieses Koeffizienten grundsätzlich zunächst ein Streudiagramm erstellt werden. Warum? Wie wir bereits angedeutet haben, ist der Produkt-Moment-Korrelationskoeffizient ein Maß für den *linearen* Zusammenhang zwischen zwei Variablen. Ist die Form des Zusammenhangs tatsächlich kurvilinear, so erbringt der Korrelationskoeffizient irreführende Ergebnisse: Er würde einen zu schwachen oder gar keinen Zusammenhang ausweisen. Eine „blinde" und automatische Berechnung des Korrelationskoeffizienten hat daher wenig Sinn. Wenn uns ein Blick auf das Streudiagramm zeigt, dass zwischen den untersuchten Variablen eine kurvilineare Beziehung besteht, sollten wir daher darauf verzichten, die Analyse in der üblichen Weise fortzuführen und den Korrelationskoeffizienten zu berechnen.[1]

Ein weiterer Grund dafür, zuerst stets ein Streudiagramm zu betrachten, besteht darin, dass Ausreißer den Korrelationskoeffizienten verfälschen können. Mit Ausreißern sind hier Messwerte einzelner Personen gemeint, die auf einer oder beiden Variablen ungewöhnlich hoch oder niedrig sind (siehe Kapitel 6). Im Streudiagramm lassen sich solche Ausreißer leicht erkennen, da die entsprechenden Datenpunkte weit von den übrigen Werten in der Punktwolke entfernt liegen. Wie genau Ausreißer den Korrelationskoeffizienten beeinflussen können und wie man mit ihnen bei der Analyse von Zusammenhängen verfahren kann, werden wir uns im Abschnitt 7.4.1 näher ansehen.

7.3 Der Produkt-Moment-Korrelationskoeffizient

Wie lässt sich nun ein Maß entwickeln, dass die Stärke und Richtung des Zusammenhangs zwischen zwei intervallskalierten Variablen in einer Zahl ausdrückt? Wie wir in den vorangegangenen Abschnitten gesehen haben, gehen bei linearen Korrelationen hohe Werte auf einer Variablen systematisch mit hohen oder niedrigen Werten auf der anderen Variablen einher. Um ein Maß für den Zusammenhang zu bestimmen, ist also zunächst zu klären, was überhaupt genau mit „hohen“ bzw. „niedrigen“ Messwerten gemeint ist. Die Antwort ist sehr simpel: Als „hoch“ gelten Messwerte, die oberhalb des Mittelwerts der entsprechenden Variable liegen. Niedrige Messwerte liegen dagegen unterhalb des Mittelwerts. Der erste Schritt bei der Berechnung des Korrelationskoeffizienten besteht daher darin, für jede Person auf beiden Variablen die Differenz zwischen ihrem Messwert und dem Mittelwert zu bestimmen. Man berechnet also die Abweichungswerte $(x_i - \bar{x})$ und $(y_i - \bar{y})$. Bei hohen Messwerten sind diese Abweichungswerte offensichtlich positiv, bei niedrigen Werten dagegen negativ. Im nächsten Schritt wird nun für jede Person das so genannte Kreuzprodukt aus ihren beiden Abweichungswerten gebildet. Hat eine Person auf beiden Variablen hohe Werte so ist natürlich auch das Kreuzprodukt $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ positiv. Ein positiver Wert resultiert aber auch dann, wenn eine Person auf beiden Variablen niedrige Messwerte aufweist – da das Produkt aus zwei negativen Abweichungswerten positiv wird.

▶Abbildung 7.6 zeigt noch einmal die Daten aus unserer fiktiven Untersuchung zum Zusammenhang zwischen dem „Konsum Gewalt darstellender Fernsehsendungen“ und der „Aggressivität“ bei Kindern. Allerdings sind nun zusätzlich die Mittelwerte beider Variablen eingezeichnet. Dadurch entstehen im Streudiagramm Quadranten. Es ist leicht erkennbar, dass bei einem positiven Zusammenhang – wie er hier gegeben ist – die meisten Datenpunkte in den Quadranten A und C liegen. Dies sind

1 Natürlich gibt es auch bei kurvilinearen Beziehungen Möglichkeiten, ein genaues Maß für die Stärke des Zusammenhangs zu bestimmen. Erläuterungen zu Verfahrensweisen, die in diesem Fall angemessen sind, findet man z.B. bei Cohen, Cohen und West (2003). Eine weitere Möglichkeit besteht darin, kurvilineare Zusammenhänge in lineare zu transformieren und diese mit den hier behandelten Verfahren für lineare Zusammenhänge weiter zu verarbeiten. Weiteres dazu erfahren Sie in Kapitel 20.

die Quadranten, in denen positive Kreuzprodukte entstehen, denn die Abweichungswerte einer Person sind hier entweder beide positiv (Quadrant A) oder beide negativ (Quadrant C). Was passiert, wenn wir die Kreuzprodukte aller Personen aufaddieren? Die Summe

$$\sum_{i}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$$

wird offensichtlich groß und positiv. Diese Summe der Kreuzprodukte wird auch Produkt-Moment genannt. Sie bildet bereits die wesentliche Grundlage für unser Zusammenhangsmaß. Dies zeigt sich auch darin, dass sich aus ihr die Bezeichnung „Produkt-Moment-Korrelationskoeffizient" ableitet.[2]

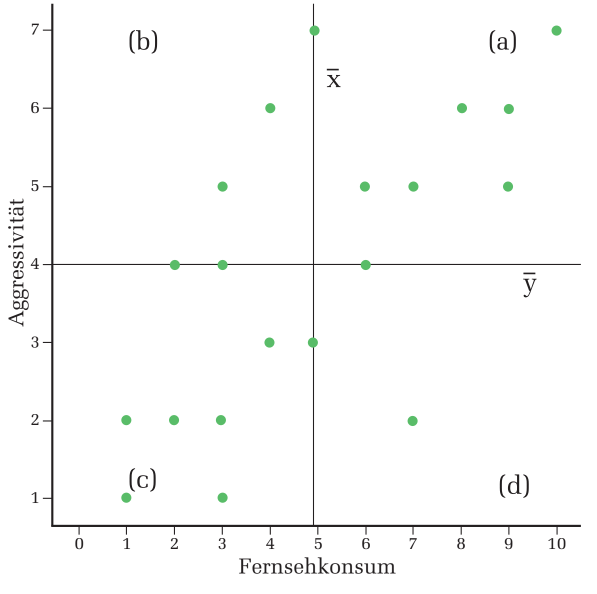

Abbildung 7.6: Das Streudiagramm zeigt dieselben Daten wie das Streudiagramm in Abbildung 7.1. Zusätzlich sind die Mittelwerte der Variablen Fernsehkonsum ($\bar{x}$) und Aggressivität ($\bar{y}$) eingezeichnet.

Überlegen wir uns kurz, wie sich das Produkt-Moment verhalten würde, wenn zwischen dem Fernsehkonsum und der Aggressivität ein negativer Zusammenhang bestünde. Bei negativen Korrelationen sind hohe Werte auf einer Variablen mit niedrigen Werten auf der anderen Variablen verbunden. Im Streudiagramm würden die meisten Datenpunkte in diesem Fall also in den Quadranten b und d liegen. Im Quadranten b befinden sich Personen, die auf der X-Variable niedrige und der Y-Variable hohe Werte aufweisen. Das Kreuzprodukt wird für diese Personen negativ, da wir jeweils einen negativen Abweichungswert mit einem positiven Abweichungswert multiplizieren. Per-

2 Die Bezeichnung „Produkt-Moment" geht auf Karl Pearson zurück. Pearson entwickelte (gemeinsam mit Francis Galton) auch die übrigen Formeln, die zur Berechnung des Produkt-Moment-Korrelationskoeffizienten notwendig sind. Dieser Koeffizient wird daher auch „Pearson-Korrelationskoeffizient" genannt.

sonen im Quadranten d haben einen hohen Wert auf der X-Variablen und einen niedrigen Wert auf der Y-Variablen. Auch hier ergeben sich also negative Kreuzprodukte, da wir jeweils einen positiven und einen negativen Wert multiplizieren. Wenn wir nun wiederum die Kreuzprodukte *aller* Personen aufaddieren, erhalten wir natürlich auch eine negative Summe.

Besteht zwischen zwei Variablen kein Zusammenhang, so verteilen sich die Datenpunkte im Streudiagramm annähernd gleichmäßig auf die Quadranten. In diesem Fall werden wir einige Personen finden, die auf beiden Variablen hohe Werte aufweisen, und solche, die auf beiden Variablen niedrige Werte haben. Für diese Personen resultieren positive Kreuzprodukte. Es wird aber ähnlich viele Personen geben, die auf einer Variablen einen hohen und auf der anderen Variablen einen niedrigen Wert erzielt haben. Bei diesen Personen ergeben sich negative Kreuzprodukte. Wenn wir nun alle Kreuzprodukte aufaddieren, so werden sich positive und negative Werte aufheben und das Produkt-Moment wird (zumindest in etwa) Null betragen.

Das Produkt-Moment hat also schon einige Eigenschaften, die für ein Zusammenhangsmaß wünschenswert sind: Es wird bei positiven Zusammenhängen positiv und bei negativen Zusammenhängen negativ. Besteht zwischen zwei Variablen kein Zusammenhang, so nimmt es den Wert 0 an. Dennoch kann das Produkt-Moment noch nicht als Zusammenhangsmaß verwendet werden. Ein Problem besteht darin, dass das Produkt-Moment von der Anzahl der Personen abhängig ist, die in die Berechnung einbezogen wurden. Hätten wir etwa unsere Studie zum Zusammenhang von Fernsehkonsum und Aggressivität mit 50 statt mit 20 Kindern durchgeführt, so wäre natürlich auch die Summe der Kreuzprodukte größer ausgefallen. Allein aufgrund des Produkt-Moments lässt sich also nicht angeben, ob ein Zusammenhang stark oder schwach ausgeprägt ist. Dieses Problem lässt sich allerdings leicht lösen: Wir dividieren das Produkt-Moment durch die Anzahl der Personen. Auf diese Weise errechnen wir das durchschnittliche Kreuzprodukt, das auch als *Kovarianz* bezeichnet wird:

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$$

Auch die Kovarianz ist als Zusammenhangsmaß noch nicht sonderlich gut geeignet. Ihr Problem besteht darin, dass sie von den Maßeinheiten abhängig ist, in denen die Variablen gemessen wurden. Anders ausgedrückt: Die Kovarianz ist nicht invariant gegenüber linearen Transformationen der Variablen (siehe Kapitel 3). Wenn wir etwa die Korrelation zwischen Körpergröße und Körpergewicht ermitteln wollten, so würde die Kovarianz höher ausfallen, wenn wir die Größe statt in Metern in Zentimetern messen. Machen wir uns dies kurz an einem Beispiel klar: Nehmen wir an, in einer Stichprobe von Personen liegt der Mittelwert der Körpergröße bei 1,75 m und der Mittelwert des Körpergewichts bei 70 kg. Eine bestimmte Person in der Stichprobe ist 1,80 m groß und 75 kg schwer. Das Kreuzprodukt dieser Person beträgt dann $(1{,}80 - 1{,}75) \cdot (75 - 70) = 0{,}25$. Wie ändert sich das Kreuzprodukt, wenn wir die Messwerte der Körpergröße mit 100 multiplizieren, also von Meter in Zentimeter umrechnen? Der Mittelwert der Größe ist dann natürlich 175 cm und das Kreuzprodukt errechnet sich als: $(180 - 175) \cdot (75 - 70) = 25$. Durch die Umrechnung in Zentimeter wird also auch das Kreuzprodukt dieser Person 100 Mal größer! Da dasselbe natürlich auch bei allen anderen Personen in der Stichprobe passiert, wächst auch die Kovarianz um den Faktor 100, wenn wir die Körpergröße in Zentimetern statt in Metern messen. Einen ähnlichen Effekt würden wir auch erzielen, wenn wir als Maßeinheit für das Gewicht Gramm anstelle von Kilogramm verwenden würden. In diesem Fall würde die Kovarianz um den Faktor 1000 wachsen. Generell gilt also für die Kovarianz: Multiplizieren wir die X-Variable mit a und die Y-Variable mit b, so wächst die Kovarianz um den Faktor $a \cdot b$.

Der „tatsächliche" Zusammenhang zwischen der Körpergröße und dem Körpergewicht ändert sich natürlich nicht in Abhängigkeit davon, ob die Größe in Metern oder in Zentimetern gemessen wird – die Information in den Messwerten ist ja immer dieselbe. Das Gleiche gilt offensichtlich auch in unserer Beispieluntersuchung: Der „tatsächliche" Zusammenhang zwischen dem Fernsehkonsum und der Aggressivität ist unabhängig davon, ob wir die Aggressivitätsunterschiede zwischen Personen auf einer Skala ausdrücken, die von 0 bis 7 oder von 0 bis 70 reicht. (In diesem Fall erfolgt die Wahl der Größe der Skala – wie meist in der Psychologie – auch mehr oder weniger willkürlich, da es keine etablierte Maßeinheit für Aggressivität gibt.) Ein geeignetes Zusammenhangsmaß sollte daher ebenfalls unabhängig von den Maßeinheiten sein, in denen die Variablen erfasst werden. Wie lässt sich dies erreichen? Die Lösung besteht darin, die Kovarianz durch das Produkt der Standardabweichungen der beiden Variablen zu teilen. Das Resultat dieser Division ist der Produkt-Moment-Korrelationskoeffizient:

$$r_{xy} = \frac{\frac{1}{n}\sum_{i}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{s_x \cdot s_y}$$

Eine Transformation der Maßeinheit wirkt auf die Standardabweichung einer Variablen genau so wie sie auf die Kovarianz wirkt. Messen wir also die Körpergröße in Zentimetern statt in Metern, so wird nicht nur die Kovarianz mit dem Körpergewicht 100 Mal größer, sondern auch die Standardabweichung der Körpergröße. (Probieren Sie dies aus! Sie können dazu die Werte aus dem Zahlenbeispiel im Abschnitt 6.4.1 verwenden. Multiplizieren Sie die Werte mit irgendeinem Faktor a und berechnen Sie die Standardabweichung der neuen Werte. Sie werden feststellen, dass die Standardabweichung ebenfalls um den Faktor a wächst.) Für den Produkt-Moment-Korrelationskoeffizienten bedeutet dies, dass jede Veränderung der Maßeinheiten zu proportionalen Änderungen im Zähler und im Nenner führt. Diese Änderungen kürzen sich also heraus. Der Korrelationskoeffizient hängt somit ausschließlich von der Stärke des linearen Zusammenhangs zwischen zwei Variablen ab und wird durch die Maßeinheiten der Variablen nicht beeinflusst.

Etwas allgemeiner können wir feststellen, dass der Produkt-Moment-Korrelationskoeffizient gegenüber linearen Transformationen der Variablen invariant ist. Sie erinnern sich: Dies waren Transformation der Form $ax + b$. Ein Beispiel für eine solche Transformation war die Umrechnung von Temperaturangaben in Celsius in Angaben in Fahrenheit (siehe Kapitel 3). Die Multiplikation mit a verändert dabei die Maßeinheit. Durch die Addition von b wird der Nullpunkt der Temperaturskala verändert (0° C entsprechen

32° F). Auch diese Veränderung des Nullpunkts hat keine Wirkung auf den Korrelationskoeffizienten. Der Grund liegt darin, dass ein veränderter Nullpunkt weder die Standardabweichung einer Variablen noch ihre Kovarianz mit einer anderen Variablen beeinflusst. Praktisch bedeutet dies z.B., dass wir für den Zusammenhang zwischen der Tageshöchsttemperatur in München und der Tageshöchsttemperatur in Hamburg immer den gleichen Produkt-Moment-Korrelationskoeffizienten ermitteln werden – unabhängig davon, ob wir die Temperatur auf der Celsius- oder der Fahrenheit-Skala messen.

Im Rechenbeispiel 7.1 wird die Berechnung des Produkt-Moment Korrelationskoeffizienten demonstriert. Das Beispiel bezieht sich auf die (fiktiven) Daten aus unserer Untersuchung zum Zusammenhang zwischen dem „Konsum gewaltdarstellender Fernsehsendungen“ und der „Aggressivität“. Die Berechnung zeigt, dass zwischen diesen Variablen eine Korrelation von $r = .65$ besteht. Was genau bedeutet dieses Ergebnis? Wir haben schon gesehen, dass der Produkt-Moment Korrelationskoeffizient bei einem perfekten negativen Zusammenhang den Wert -1 und bei einem perfekten positiven Zusammenhang den Wert 1 annimmt. Zwischen dem Fernsehkonsum und der Aggressivität besteht also ein deutlich ausgeprägter, positiver Zusammenhang, der aber andererseits auch recht weit davon entfernt ist, perfekt zu sein. Dennoch würde eine Korrelation von $r = .65$ in der Psychologie als starker Zusammenhang beurteilt werden. Im Kasten „Was ist eine große Korrelation?“ wird erläutert, warum dies so ist.

Rechenbeispiel 7.1 *Zusammenhang zwischen Aggressivität und Fernsehkonsum* ▶

▶

▶Fortsetzung

Teilnehmer-Nr.	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\cdot(y_i-\bar{y})$
1	3	4	−1,9	0	3,61	0	0
2	1	1	−3,9	−3	15,21	9	11,7
3	2	2	−2,9	−2	8,41	4	5,8
4	2	4	−2,9	0	8,41	0	0
5	4	3	−0,9	−1	0,81	1	0,9
6	3	2	−1,9	−2	3,61	4	3,8
7	6	5	1,1	1	1,21	1	1,1
8	6	4	1,1	0	1,21	0	0
9	7	2	2,1	−2	4,41	4	−4,2
10	8	6	3,1	2	9,61	4	6,2
11	9	6	4,1	2	16,81	4	8,2
12	9	5	4,1	1	16,81	1	4,1
13	5	3	0,1	−1	0,01	1	−0,1
14	10	7	5,1	3	26,01	9	15,3
15	3	1	−1,9	−3	3,61	9	5,7
16	5	7	0,1	3	0,01	9	0,3
17	3	5	−1,9	1	3,61	1	−1,9
18	1	2	−3,9	−2	15,21	4	7,8
19	4	6	−0,9	2	0,81	4	−1,8
20	7	5	2,1	1	4,41	1	2,1
	$\Sigma = 98$	$\Sigma = 80$	$\Sigma = 0$	$\Sigma = 0$	$\Sigma = 143{,}8$	$\Sigma = 70$	$\Sigma = 65$
	$\bar{x} = 4{,}9$	$\bar{y} = 4$			$s_x^2 = \frac{143{,}8}{20} = 7{,}19$ $s_x = 2{,}68$	$s_y^2 = \frac{70}{20} = 3{,}5$ $s_y = 1{,}87$	$\operatorname{cov}(x,y) = \frac{65}{20}$ $= 3{,}25$

Tabelle 7.1: Berechnung des Produkt-Moment-Korrelationskoeffizienten in der Studie zum Zusammenhang zwischen Fernsehkonsum und Aggressivität (fiktive Daten).

HINTERGRUND

Was ist eine große Korrelation? In der Psychologie werden oftmals Konventionen verwendet, um die Stärke des Zusammenhangs zwischen zwei Variablen zu beurteilen. Diese Konventionen wurden ursprünglich von Cohen (1988) vorgeschlagen. Demnach gilt:

- $|r| \approx 0{,}1$: schwacher Zusammenhang
- $|r| \approx 0{,}3$: mittlerer Zusammenhang
- $|r| \approx 0{,}5$: starker Zusammenhang

Cohens Konventionen haben sich in der psychologischen Forschungspraxis bewährt. In vielen Gebieten der Psychologie entsprechen Zusammenhänge „mittlerer Stärke" im Bereich von $r = 0{,}3$ recht gut dem Durchschnitt der Korrelationskoeffizienten, die tatsächlich gefunden werden. Korrelationen im Bereich von $r = 0{,}5$ (oder noch größere Korrelationen) treten dagegen selten auf und werden entsprechend als starke Zusammenhänge bewertet.

Nun ist eine Korrelation von $r = 0{,}5$ von einem perfekten Zusammenhang sehr weit entfernt. (Dies illustriert auch das Streudiagramm in Abbildung 7.5d, in dem eine Korrelation von $r = -0{,}45$ dargestellt ist.) Warum sind höhere Korrelationen in der Psychologie dennoch selten? Ein wesentlicher Grund besteht darin, dass die meisten psychologischen Variablen von einer ganzen Vielzahl anderer Variablen beeinflusst werden. Nehmen wir für einen Moment an, dass das häufige Betrachten Gewalt darstellender Fernsehsendungen bei Kindern tatsächlich höhere Aggressivität verursacht (dies wird auch durch eine starke Korrelation zwischen Fernsehkonsum und Aggressivität *nicht* belegt – siehe Abschnitt 7.5). Sollten wir dann erwarten, dass zwischen diesen Variablen ein perfekter Zusammenhang besteht? Nein, denn damit würden wir übersehen, dass wahrscheinlich noch eine ganze Reihe weiterer Variablen die Aggressivität von Kindern beeinflusst. So können wir beispielsweise mit guten Gründen vermuten, dass auch der Erziehungsstil der Eltern und die Häufigkeit, mit der ein Kind frustrierende Erfahrungen macht, auf die Aggressivität wirken. Diese und alle weiteren Faktoren, die die Aggressivität möglicherweise beeinflussen, werden in der Korrelation zwischen dem Fernsehkonsum und der Aggressivität nicht berücksichtigt. Es wäre daher naiv zu erwarten, dass wir allein aufgrund der Kenntnis des Fernsehkonsums eines Kindes präzise und fehlerfrei angeben können, wie aggressiv es sich verhält. Die meisten psychologischen Variablen stehen – wie die Aggressivität – in Beziehung zu zahlreichen anderen Variablen. Eine einzelne Korrelation, die lediglich den Zusammenhang zwischen zwei Variablen zum Ausdruck bringt, ist ungeeignet, um dieses Beziehungsgeflecht vollständig abzubilden. Es gibt also gute Gründe, Cohens Konventionen zu folgen und Korrelationen im Bereich von $r = 0{,}5$ bereits als starke Zusammenhänge zu bewerten. Dabei ist allerdings zu beachten, dass die Konventionen lediglich eine Orientierungshilfe darstellen. Diese Orientierungshilfe ist insbesondere dann nützlich, wenn Korrelationen auf einem neuen Forschungsgebiet, zu dem noch wenig gesichertes Wissen vorhanden ist, zu beurteilen sind. Ist es jedoch möglich, einen Zusammenhang zwischen zwei Variablen sinnvoll mit anderen bereits bekannten Korrelationen zu vergleichen, so liefern ausschließlich diese anderen Korrelationen den relevanten Bewertungsmaßstab: Nehmen wir an, dass bereits bekannt ist, dass von 20 theoretisch plausiblen Variablen keine stärker als mit $r = 0{,}15$ mit der Arbeitszufriedenheit von Angestellten korreliert.

▸Fortsetzung

Wird dann entdeckt, dass eine weitere Variable eine Korrelation von $r = 0{,}30$ mit der Arbeitszufriedenheit aufweist, so ist diese Korrelation in diesem Forschungsfeld natürlich als ungewöhnlich starker Zusammenhang zu bewerten – und nicht als mittlerer Zusammenhang. Andererseits können sich auch vermeintlich hohe Korrelationen im Kontext relevanter Forschungsergebnisse als niedrig erweisen: Existieren diverse diagnostische Verfahren, deren Ergebnisse mit $r = 0{,}6$ mit der Berufseignung zum Schlosser korrelieren, so ist ein alternatives Verfahren, dessen Ergebnisse mit $r = 0{,}45$ mit der Berufseignung korrelieren, als schlecht zu beurteilen. (Einige weitere Hinweise zur Bewertung und Interpretation von Korrelationen finden Sie im Abschnitt 9.5.)

7.3.1 z-Werte und der Produkt-Moment-Korrelationskoeffizient

Der Produkt-Moment-Korrelationskoeffizient zwischen zwei Variablen kann auch mit Hilfe von z-Werten berechnet werden. Die z-Standardisierung haben wir bereits im Kapitel 6 kennen gelernt. Fassen wir noch einmal zusammen: z-Werte ermöglichen es unter anderem, die Messwerte einer Person auf zwei unterschiedlichen Variablen miteinander zu vergleichen. Dies wird möglich, da z-Werte die Größe eines Messwerts nicht in den ursprünglichen Maßeinheiten ausdrücken. Sie geben stattdessen die Lage eines Messwerts relativ zu den übrigen Werten in einer Stichprobe an: Ein positiver z-Wert zeigt an, dass der entsprechende Messwert oberhalb des Mittelwerts liegt. Ein negativer z-Wert besagt dagegen, dass der Messwert kleiner ist als der Mittelwert. Hat eine Person auf zwei Variablen jeweils einen z-Wert von 0, so bedeutet dies, dass sie auf beiden Variablen eine exakt durchschnittliche Ausprägung aufweist. Den Abstand eines Messwerts vom Mittelwert geben z-Werte in Standardabweichungen an. Ein z-Wert von 1,5 heißt also nichts anderes, als dass der Messwert 1,5 Standardabweichungen größer ist als der Mittelwert. Ein z-Wert von -2 zeigt dagegen an, dass der Messwert 2 Standardabweichungen unterhalb des Mittelwerts liegt.

Schon diese kurze Zusammenfassung macht deutlich, dass z-Werte bei der Berechnung des Produkt-Moment Korrelationskoeffizienten hilfreich sein können: Auch bei der Berechnung des Zusammenhangsmaßes haben wir in einem ersten Schritt die Abweichung der Messwerte von ihrem Mittelwert bestimmt – allerdings zunächst in den ursprünglichen Maßeinheiten. Die Unabhängigkeit des Korrelationskoeffizienten von den Maßeinheiten wurde schließlich aber mit einer Division durch die Standardabweichungen erreicht.

Wie lässt sich der Produkt-Moment Korrelationskoeffizient nun genau mit Hilfe von z-Werten errechnen? Dies wird leicht erkennbar, wenn wir die Formel des Korrelationskoeffizienten umstellen:

$$r_{xy} = \frac{\frac{1}{n}\sum_i^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{s_x \cdot s_y} = \frac{1}{n}\sum_i^n \frac{(x_i - \bar{x}) \cdot (y_i - \bar{y})}{s_x \cdot s_y} = \frac{1}{n}\sum_i^n \frac{(x_i - \bar{x})}{s_x} \cdot \frac{(y_i - \bar{y})}{s_y} = \frac{1}{n}\sum_i^n z_x \cdot z_y$$

Der Korrelationskoeffizient entspricht also dem durchschnittlichen Kreuzprodukt (bzw. der Kovarianz) der z-Werte.

Wie die Umstellung der Formel zeigt, nimmt der Korrelationskoeffizient dann den Wert 1 an, wenn die Summe der Kreuzprodukte der z-Werte genau so groß ist wie die Anzahl n der Personen in der Stichprobe. Dies ist aber nur dann der Fall, wenn jede Person auf der X-Variablen exakt den gleichen z-Wert erhält wie auf der Y-Variablen. Bei einem perfekten Zusammenhang hat also jede Person in einer Stichprobe auf beiden Variablen exakt die gleiche relative Position. Dies lässt sich auch verallgemeinern: Hat jede Person auf beiden Variablen zumindest eine ähnliche relative Position – und somit ähnliche z-Werte, so resultiert ein hoher Korrelationskoeffizient. Je deutlicher sich die z-Werte auf beiden Variablen unterscheiden, desto geringer wird der Korrelationskoeffizient. Generell wird durch den Produkt-Moment Korrelationskoeffizienten also ausgedrückt, in welchem Ausmaß Personen auf zwei Variablen die gleiche relative Position einnehmen.

Im Rechenbeispiel 7.2 wird die Berechnung des Korrelationskoeffizienten mit z-Werten an den Daten aus unserer Studie zu „Fernsehkonsum und Aggressivität“ illustriert.

Rechenbeispiel 7.2 *Zusammenhang zwischen Aggressivität und Fernsehkonsum in z-Werten*

▶Fortsetzung

Teilnehmer-Nr.	x_i	y_i	z_x	z_y	$z_x \cdot z_y$
1	3	4	–0,71	0	0
2	1	1	–1,46	–1,60	2,34
3	2	2	–1,08	–1,07	1,16
4	2	4	–1,08	0	0
5	4	3	–0,34	–0,53	0,18
6	3	2	–0,71	–1,07	0,76
7	6	5	0,41	0,53	0,22
8	6	4	0,41	0	0
9	7	2	0,78	–1,07	–0,83
10	8	6	1,16	1,07	1,24
11	9	6	1,53	1,07	1,64
12	9	5	1,53	0,53	0,81
13	5	3	0,04	–0,53	–0,02
14	10	7	1,90	1,60	3,04
15	3	1	–0,71	–1,60	1,14
16	5	7	0,04	1,60	0,06
17	3	5	–0,71	0,53	–0,38
18	1	2	–1,46	–1,07	1,56
19	4	6	–0,34	1,07	–0,36
20	7	5	0,78	0,53	0,41
			$\Sigma = 0$	$\Sigma = 0$	$\Sigma = 12{,}97$

Tabelle 7.2: Berechnung des Produkt-Moment-Korrelationskoeffizienten in der Studie zum Zusammenhang zwischen Fernsehkonsum und Aggressivität (fiktive Daten) über *z*-Werte.

7.4 Verzerrungen des Produkt-Moment-Korrelationskoeffizienten

Der Produkt-Moment-Korrelationskoeffizient wird durch eine Reihe von Faktoren beeinflusst. Unter Umständen können diese Faktoren dazu führen, dass der Korrelationskoeffizient die tatsächliche Stärke des Zusammenhangs zwischen zwei Variablen *nicht* korrekt wiedergibt. Es kommt also zu Verzerrungen des Zusammenhangsmaßes. Bei der Interpretation des Produkt-Moment-Korrelationskoeffizienten sollte natürlich darauf geachtet werden, ob solche Verzerrungen eingetreten sind.

7.4.1 Ausreißerwerte

Messwerte einzelner Personen, die auf einer oder beiden Variablen deutlich höher oder niedriger sind als die Messwerte der übrigen Personen, können einen sehr starken Effekt auf den Produkt-Moment-Korrelationskoeffizienten haben. Solche Ausreißerwerte gefährden die Aussagekraft des Korrelationskoeffizienten insbesondere dann, wenn ein Zusammenhang zwischen zwei Variablen lediglich an einer kleinen Stichprobe untersucht wird. Die beiden Streudiagramme in ▶Abbildung 7.7 verdeutlichen das Problem. In beiden Streudiagrammen sind Messwerte von 11 Teilnehmern dargestellt, wobei jeweils ein Datenpunkt leicht als Ausreißerwert identifiziert werden kann. Im Streudiagramm a führt dieser Ausreißer zu einer dramatischen Erhöhung des Korrelationskoeffizienten. Bei den 10 eng zusammen liegenden Punkten besteht nahezu kein Zusammenhang zwischen den Variablen X und Y ($r = 0{,}03$). Bezieht man in die Berechnung jedoch zusätzlich den Ausreißerwert ein, so ergibt sich ein Korrelationskoeffizient von $r = 0{,}90$! Das Streudiagramm b illustriert, dass Ausreißer den Korrelationskoeffizienten auch vermindern können. Hier besteht für 10 der 11 Personen ein starker Zusammenhang zwischen X und Y ($r = 0{,}90$). Dass der Korrelationskoeffizient in der gesamten Stichprobe dennoch nur $r = 0{,}29$ beträgt, geht ausschließlich auf die ungewöhnlichen Messwerte des 11. Untersuchungsteilnehmers zurück.

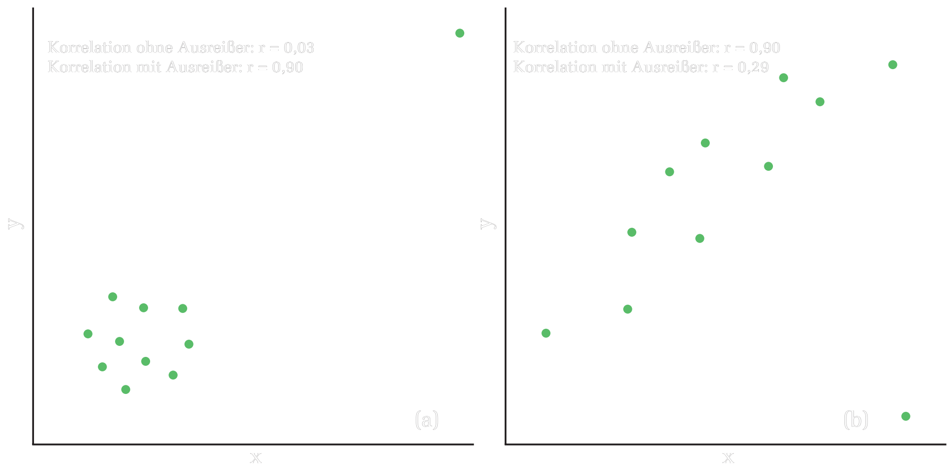

Abbildung 7.7: Illustration des Effekts von Ausreißerwerten auf den Produkt-Moment-Korrelationskoeffizienten. Im Diagramm (a) wird der Korrelationskoeffizient durch den Ausreißerwert erhöht, im Diagramm (b) führt der Ausreißer zu einer Verminderung des Korrelationskoeffizienten.

Wie sollte man bei der Interpretation des Korrelationskoeffizienten verfahren, wenn Ausreißerwerte vorliegen? Auf keinen Fall sollte man solche Ausreißer ignorieren. Bei der im Streudiagramm a in Abbildung 7.7 dargestellten Situation sollte man also nicht schließen, dass zwischen X und Y tatsächlich ein starker Zusammenhang besteht. Eine Aussage über den Zusammenhang zweier Variablen sollte natürlich nicht in derart entscheidender Weise von den Messwerten einer einzelnen Person abhängen! Letztlich hängt die weitere Vorgehensweise davon ab, welche Ursachen dazu führten, dass ein Ausreißer auftrat. Bestehen gute Gründe für die Annahme, dass die ungewöhnlichen Werte auf einen Fehler in der Messprozedur zurückgehen, so kann der Ausreißer bei der Berechnung des Korrelationskoeffizienten ausgeschlossen werden. Dies wäre zum Beispiel angemessen, wenn ein Teilnehmer bei einer Reaktionsaufgabe das Startsignal „verschlafen" hat, wenn er bei der Bearbeitung eines psychometrischen Tests auffällig unmotiviert war, wenn er alkoholisiert zu der Studie erschien oder wenn er die Instruktion eines Fragebogens nicht verstanden hat. Vermutet man hingegen, dass der ungewöhnliche Messwert die „wirkliche" Merkmalsausprägung eines Teilnehmers korrekt wiedergibt, so ist es nicht sinnvoll (und auch nicht zulässig), den Ausreißer von der weiteren Auswertung auszuschließen. In diesem Fall besteht die beste Möglichkeit wohl darin, den Zusammenhang zwischen den Variablen X und Y an einer größeren Stichprobe erneut zu untersuchen. Unabhängig davon, für welche Verfahrensweise man sich entscheidet, sollte man bei einem späteren Bericht natürlich in jedem Fall angeben, dass Ausreißer aufgetreten sind und wie man mit ihnen bei der Analyse umgegangen ist.

7.4.2 Einschränkungen der Variabilität

Nehmen wir an, dass in einer Studie die Beziehung zwischen dem Lebensalter und den Mathematikkenntnissen bei Kindern zwischen 6 und 16 Jahren untersucht wird. Die Studie würde wahrscheinlich zu dem wenig überraschenden Ergebnis führen, dass zwischen dem Lebensalter und den Mathematikkenntnissen ein starker Zusammenhang besteht. Würde die Studie aber das gleiche Resultat erbringen, wenn – um Zeit und Aufwand zu sparen – lediglich Kinder zwischen 10 und 13 Jahren in die Untersuchung einbezogen werden? Die Antwort auf diese Frage lautet „Nein": In diesem Fall würde ein deutlich geringerer Zusammenhang zwischen dem Lebensalter und den Mathematikkenntnissen gefunden werden. Warum?

Das Problem besteht hier darin, dass durch die Begrenzung auf Kinder aus einem bestimmten Altersbereich, die Variabilität des Merkmals Lebensalters eingeschränkt wird: Viele mögliche Werte dieses Merkmals – die für eine allgemeine Aussage über den Zusammenhang zwischen dem Alter und den Mathematikkenntnissen bei Kindern durchaus interessant wären – können in unserer Stichprobe gar nicht mehr auftreten. Solche Variabilitätseinschränkungen führen generell zu einer Verminderung des Korrelationskoeffizienten. Kennen wir also die Korrelation zwischen dem Lebensalter und den Mathematikkenntnissen bei Kindern zwischen 10 und 13 Jahren, so können wir noch nicht angeben, wie stark dieser Zusammenhang bei Kindern zwischen 6 und 16 Jahren ist. Ebenso erlaubt uns der Korrelationskoeffizient in unserer Studie mit Kindern zwischen 6 und 16 Jahren noch keine Aussage über die Stärke des Zusammenhangs bei Kindern zwischen 3 und 16 Jahren.

Die beiden Streudiagramme in ▶Abbildung 7.8 verdeutlichen, wie es zu dieser Verringerung des Korrelationskoeffizienten kommt. Das Streudiagramm a in Abbildung 7.8 zeigt zunächst die vollständigen Ergebnisse der Studie mit Kindern zwischen 6 und 16 Jahren. Hier ist leicht ein eindeutiger Trend in Daten zu identifizieren. Offensichtlich besteht ein deutlicher Zusammenhang zwischen dem Lebensalter und den Mathematikkenntnissen. Im Streudiagramm b in Abbildung 7.8 ist der Ausschnitt aus den Daten wiedergegeben, der in einer Studie mit Kindern zwischen 10 und 13 Jahren gefunden worden wäre. Augenscheinlich besteht in diesem Altersbereich nur eine geringe Korrelation zwischen den Variablen. Dies ist darauf zurückzuführen, dass der gesamte Anstieg in den Mathematikkenntnissen, der in den Punktwolken erkennbar wird, im Altersbereich von 10 bis 13 Jahren natürlich geringer ausfällt als im Altersbereich von 6 bis 16 Jahren. Würden wir in die Streudiagramme wiederum die Gerade einzeichnen, die den allgemeinen Trend in den Daten am besten beschreibt, so würde diese Gerade zwischen dem 6. und 16. Lebensjahr von einem Mathematikkenntnis-Wert von etwa 24,5 auf einen Wert von etwa 33,5 ansteigen. Im Streudiagramm b würde die Gerade hingegen lediglich von einem Mathematikkenntnis-Wert von etwa 29 im 10. Lebensjahr auf einen Wert von etwa 31,5 im 13. Lebensjahr ansteigen. Dieser geringere „Gesamt-Anstieg" im Altersbereich von 10 bis 13 Jahren ist in der Punktwolke natürlich schwerer auszumachen – er ist in Relation zu der Abweichung der einzelnen Punkte von der Gerade zu klein, um leicht zu erkennen zu sein. Damit besteht in diesem Altersbereich auch nur eine vergleichsweise geringe Korrelation.

Wir haben zu Beginn dieses Kapitels erläutert, dass der Zusammenhang zwischen zwei Variablen um so stärker ist, je näher die Datenpunkte an der Geraden liegen, die den Trend in den Daten am besten beschreibt. Wie jetzt deutlich wird, war diese Aussage nicht ganz korrekt – oder zumindest unvollständig. Für die Stärke eines Zusammenhangs ist letztlich das Verhältnis zwischen der Abweichung der Datenpunkte und dem gesamten Anstieg der Geraden im betrachteten Wertebereich entscheidend.

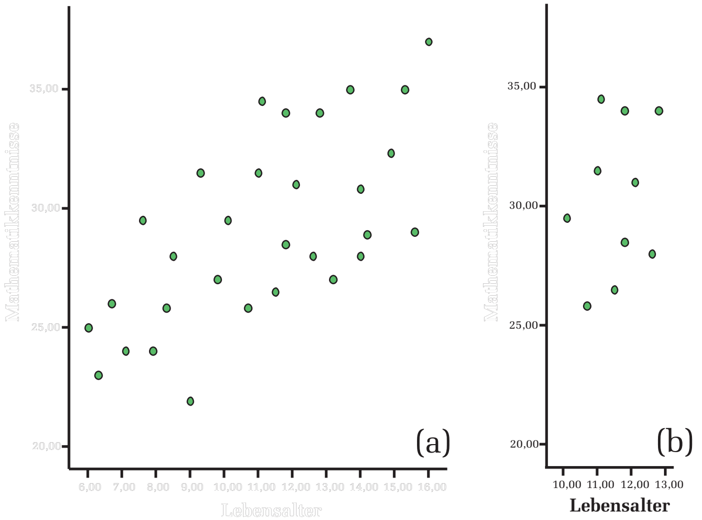

Abbildung 7.8: Ein Beispiel für den Effekt einer Variabilitätseinschränkung auf den Produkt-Moment Korrelationskoeffizienten. Im Diagramm (b) ist der Wertebereich des Lebensalters im Vergleich zum Diagramm (a) eingeschränkt.

Es ist nicht selten, dass die Bedeutung von Variabilitätseinschränkungen bei der Interpretation des Korrelationskoeffizienten übersehen wird. Betrachten wir zwei Beispiele: Zuweilen evaluieren Unternehmen die Nützlichkeit von Testverfahren, die sie bei der Personalauswahl einsetzen, indem sie die Testergebnisse mit der beruflichen Leistung korrelieren. Nun lässt sich die berufliche Leistung in der Regel nur bei Personen feststellen, die tatsächlich eingestellt wurden. Eingestellt werden aber nur solche Personen, die ein gutes Testergebnis erzielt haben. Hinsichtlich des Merkmals „Testergebnis“ liegt also eine Variabilitätseinschränkung vor. Die Korrelation zwischen den Testergebnissen und der beruflichen Leistung in der Gruppe *aller Bewerber* wird auf diese Weise unterschätzt werden. Dies ist aber die Korrelation, die für die Unternehmen eigentlich von Interesse wäre!

Das zweite Beispiel ist eng mit der gängigen psychologischen Forschungspraxis verbunden. Ein Großteil der psychologischen Forschung findet an Universitäten statt. Dabei nehmen an vielen Studien ausschließlich Studierende teil. Bei der Untersuchung von Korrelationen ist dies dann problematisch, wenn es um Merkmale geht, bei denen sich die Messwerte von Studierenden nicht über den gesamten möglichen Wertebereich verteilen. So treten bei Studierenden nur selten deutlich unterdurchschnittliche Intelligenzwerte auf. Korreliert man nun die Intelligenz mit irgendeinem anderen Merkmal, so resultiert wahrscheinlich ein schwacher Zusammenhang. Jedenfalls wäre eine stärkere Korrelation zu erwarten, wenn der Zusammenhang an einer für die Gesamtbevölkerung repräsentativen Stichprobe untersucht würde.

7.4.3 Zusammenfassung von heterogenen Untergruppen

Das Streudiagramm in ▶Abbildung 7.9a zeigt Daten aus einer fiktiven Studie zum Zusammenhang zwischen dem Umfang des wöchentlichen Krafttrainings und der Kraft, die die Probanden bei einer bestimmten Übung aufbringen können. Die Ergebnisse von Männern und Frauen sind dabei gesondert dargestellt. Die Messwerte männlicher Probanden werden in Form von Kreisen wiedergegeben, die Daten von Frauen sind durch Quadrate kenntlich gemacht. Augenscheinlich besteht in der Gruppe der Männer ein deutlicher Zusammenhang zwischen dem Umfang des Krafttrainings und der Kraft. Legen wir gedanklich eine Gerade durch die Menge der Kreise, so zeigt sich, dass die Datenpunkte nur geringfügig von dieser Gerade abweichen. Entsprechend erreicht der Korrelationskoeffizient bei den Männern einen hohen Wert von $r = 0{,}68$. In der Gruppe der Frauen ergibt sich ein ähnliches Bild: Auch die Quadrate weichen nur wenig von einer gedachten Geraden durch die entsprechende Punktwolke ab. Der Korrelationskoeffizient ist mit $r = 0{,}70$ nur unwesentlich höher als bei den Männern.

Was passiert aber, wenn wir den Korrelationskoeffizienten für Männer und Frauen gemeinsam berechnen? In diesem Fall liegt die Gerade, die den Trend *aller* Daten am besten beschreibt, *zwischen* den Messwerten von Männern und Frauen. Dies führt dazu, dass die einzelnen Datenpunkte von dieser Gerade deutlich weiter entfernt liegen. Entsprechend verringert sich der Wert des Korrelationskoeffizienten. Er beträgt in der Gesamtgruppe lediglich $r = 0{,}27$.

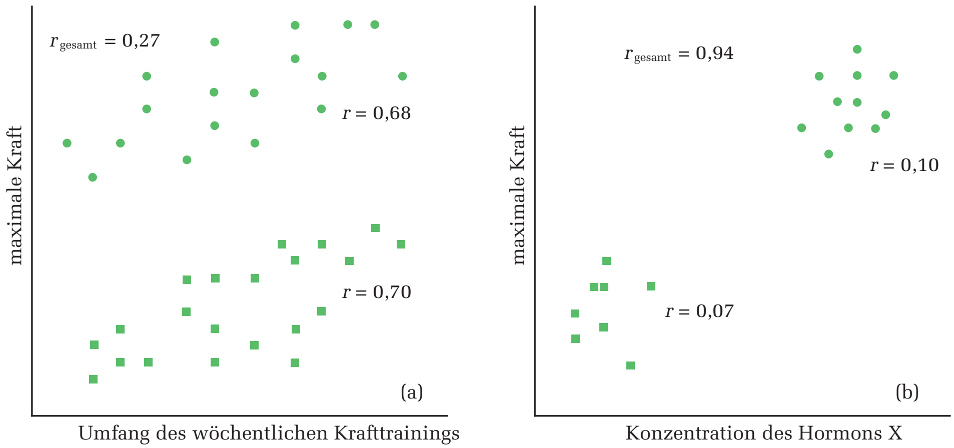

Abbildung 7.9: Illustration des Effekts, den die Zusammenfassung zweier heterogener Untergruppen auf den Produkt-Moment Korrelationskoeffizienten ausüben kann. Im Diagramm (a) ist die Korrelation in der Gesamtstichprobe deutlich niedriger als in den Untergruppen. Im Diagramm (b) steigt die Korrelation durch die Zusammenfassung der Untergruppen.

Diese Verringerung des Korrelationskoeffizienten geht darauf zurück, dass zwischen Männern und Frauen hinsichtlich der maximalen Kraft Niveauunterschiede bestehen: Bei Männern ist der Mittelwert der Kraft um einiges höher als bei Frauen. Generell können Niveauunterschiede zwischen verschiedenen Untergruppen zur Folge haben, dass der Wert des Korrelationskoeffizienten in der Gesamtgruppe deutlich von den Korrelationskoeffizienten in den Untergruppen abweicht. Das Streudiagramm in ▶Abbildung 7.9b illustriert, dass der Korrelationskoeffizient in der Gesamtgruppe dabei auch höher ausfallen kann. Das Streudiagramm könnte etwa den Zusammenhang zwischen der Konzentration eines bestimmten Hormons im Blut und der Kraft darstellen. Nehmen wir an, dass dieses Hormon bei Männern in höherer Konzentration auftritt als bei Frauen. Dann liegen sowohl auf der X-Variable (Hormonkonzentration) als auch auf der Y-Variable (Kraft) Niveauunterschiede zwischen Männern und Frauen vor. Wie das Streudiagramm zeigt, können wir in diesem Fall in der Gesamtgruppe selbst dann einen starken Zusammenhang ermitteln, wenn die Variablen Hormonkonzentration und Kraft bei Männern und Frauen nur schwach korreliert sind.

Die beiden vorangegangenen Beispiele sollten deutlich machen, dass die Daten aus heterogenen Untergruppen, zwischen denen Niveauunterschiede bestehen, nach Möglichkeit *nicht* zusammengefasst werden sollten. Stattdessen sollten gesonderte Korrelationskoeffizienten für die einzelnen Untergruppen bestimmt werden. Selbstverständlich würden wir einen Fehler machen, wenn wir aus den Daten im Streudiagramm a in Abbildung 7.9 schlössen, dass die Korrelation zwischen dem Umfang des Krafttrainings und der Kraft $r = 0{,}27$ beträgt. Der Zusammenhang zwischen Training und Trainingsergebnis ist sowohl bei Männern als auch bei Frauen deutlich stärker ausgeprägt! Ebenso wäre es falsch, aufgrund der Daten im Streudiagramm b in Abbildung 7.9 anzunehmen, dass zwischen der Hormonkonzentration und der Kraft ein starker Zusammenhang besteht. Die Hormonkonzentration unterscheidet sich zwar – genau wie die Kraft –

deutlich bei Männern und Frauen, aber in beiden Untergruppen findet sich nahezu kein Zusammenhang zwischen den Variabeln. Dies legt darüber hinaus den Schluss nahe, dass die Kraftunterschiede zwischen Männern und Frauen nicht auf die Konzentration dieses Hormons zurückgehen.

7.5 Korrelation und Kausalität

Ein recht häufiger Fehler bei der Interpretation von Korrelationen besteht in der Annahme, dass ein Zusammenhang zwischen Variablen zugleich bedeutet, dass die eine Variable die andere Variable ursächlich beeinflusst. Wir haben in diesem Kapitel bereits mehrfach darauf hingewiesen, dass eine Korrelation zwischen zwei Merkmalen keine eindeutige Aussage über eine Ursache-Wirkungs-Beziehung zwischen den Merkmalen erlaubt. Betrachten wir dieses Problem etwas genauer.

Zunächst ist es richtig, dass zwischen zwei Variablen ausschließlich dann eine Ursache-Wirkungs-Beziehung bestehen kann, wenn die Variablen auch korrelieren. Vermuten wir etwa, dass größeres Selbstbewusstsein zu größerem beruflichen Erfolg führt, so kann diese Hypothese nur korrekt sein, wenn Menschen mit größerem Selbstbewusstsein (zumindest tendenziell) auch größeren Berufserfolg haben. Besteht zwischen diesen Variablen hingegen ein Zusammenhang von $r = 0$, so ist das Selbstbewusstsein offensichtlich keine Ursache des Berufserfolgs. Eine Korrelation zwischen zwei Variablen ist also eine notwendige Voraussetzung für die Schlussfolgerung, dass die Ausprägung einer Variablen die Ausprägung der anderen Variablen ursächlich beeinflusst.

Andererseits ist eine Korrelation zwischen zwei Merkmalen aber nicht ausreichend, um auf eine Ursache-Wirkungs-Beziehung zu schließen. Generell kann eine Korrelation zwischen den Variablen X und Y auf unterschiedlichen Wegen zustande kommen. Die drei einfachsten Möglichkeiten sind in ▶Abbildung 7.10a dargestellt. Zunächst kann die Ausprägung der Variable X die Ursache der Ausprägung der Variable Y sein. Denkbar wäre aber auch, dass die Kausalrichtung anders herum verläuft: Vielleicht ist die Variable Y die Ursache und die Variable X die Folge. Schließlich kann zwischen den Variablen X und Y auch überhaupt keine Kausalbeziehung bestehen. Möglich ist nämlich auch, dass die Ausprägung beider Variablen durch eine dritte Variable Z verursacht wird.

Finden wir also zwischen dem Selbstbewusstsein und dem Berufserfolg einen nennenswerten Zusammenhang, so lässt sich dies auf mindestens drei Weisen erklären (▶Abbildung 7.10b): Vielleicht hat das Selbstbewusstsein tatsächlich einen Einfluss auf den Berufserfolg. Alternativ könnte aber auch großer Berufserfolg zu gesteigertem Selbstbewusstsein führen, während Menschen, in deren Berufsleben sich Fehlschläge aneinanderreihen, kein hohes Selbstbewusstsein entwickeln. Nicht weniger plausibel ist die dritte Möglichkeit: Ein hohes Leistungsvermögen (also eine dritte Variable) verursacht ein starkes Selbstbewusstsein und bedingt zugleich großen beruflichen Erfolg.

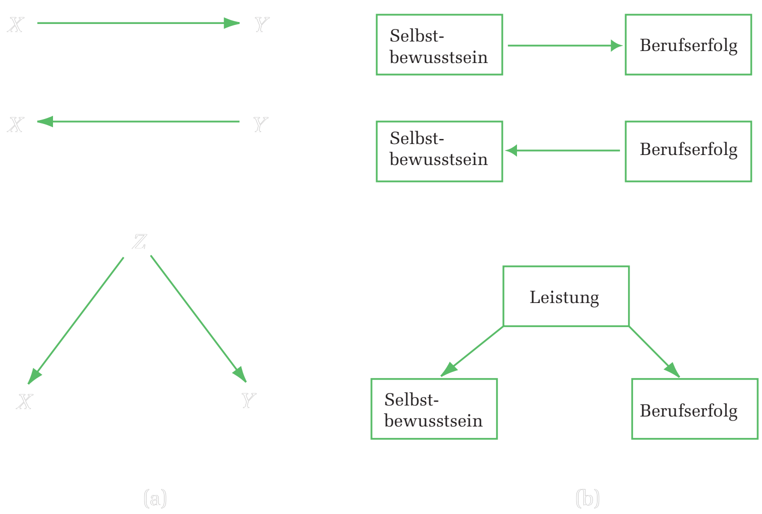

Abbildung 7.10: Kausale Beziehungen, die einem Zusammenhang zwischen zwei Variablen im Allgemeinen (a) und dem Zusammenhang zwischen Selbstbewusstsein und Berufserfolg zugrunde liegen könnten (b).

Das gleiche Erklärungsschema lässt sich natürlich auch auf den Zusammenhang zwischen dem „Konsum Gewalt darstellender Fernsehsendungen“ und der „Aggressivität“ aus unserer Beispielstudie anwenden. Es ist möglich, dass das häufige Betrachten von Gewaltdarstellungen tatsächlich dazu führt, dass sich Kinder aggressiver verhalten. Denkbar ist aber auch, dass sich aggressive Kinder einfach stärker für Gewaltdarstellungen interessieren, und folglich häufiger entsprechende Fernsehsendungen schauen. Schließlich könnte der Zusammenhang auch durch eine dritte Variable vermittelt sein. Eine solche dritte Variable könnte etwa das Ausmaß elterlicher Fürsorge sein. Vielleicht sitzen Kinder, die viel Zeit allein verbringen, länger vor dem Fernseher *und* entwickeln, aufgrund eines Mangels an elterlicher Betreuung, eine größere Aggressivität. Möglich wäre auch, dass sowohl die Neigung zu aggressivem Verhalten wie auch das Interesse an Gewaltdarstellungen bis zu einem gewissen Grade erblich bedingt sind. Es gibt vermutlich zahllose Drittvariablen, von denen sich zumindest nicht ausschließen lässt, dass sie den Zusammenhang zwischen Fernsehkonsum und Aggressivität vermitteln.

Die Versuchung, Korrelationen als Beleg für eine Kausalbeziehung zu interpretieren, hängt offensichtlich stark von den Variablen ab, zwischen denen ein Zusammenhang besteht. Im Alltag gibt es viele Beispiele für Korrelationen, aus denen wohl niemand auf eine Ursache-Wirkungs-Beziehung schließen würde. Bei Kindern zwischen 3 und 13 Jahren besteht zweifellos ein deutlicher Zusammenhang zwischen der Körpergröße und den Mathematikkenntnissen. Dennoch ist es nur zu offensichtlich, dass die Körpergröße nicht die Ursache der Mathematikkenntnisse ist. Andererseits werden Korrelationen in vielen Medien erstaunlich häufig als Ursache-Wirkungs-Beziehungen dargestellt – augenscheinlich ohne dass dies als Fehler erkannt wird. So kann man etwa in diversen Zeitschriften immer wieder Artikel mit der Empfehlung finden, mehr von einem bestimmten Lebensmittel (z.B. Olivenöl, Fisch, Rotwein) zu verzehren. Als Begründung für diese Empfehlung wird oftmals angeführt, dass die Lebenserwartung in Ländern, in denen die Einwohner viel von dem fraglichen Lebensmittel verbrauchen, höher sei. Dies ist aber ein schwaches Argument, um Ihnen anzuraten, Ihre Mahlzeiten häufiger mit Olivenöl zuzubereiten. Offensichtlich unterscheiden sich verschiedene Länder nicht nur hinsichtlich der durchschnittlichen Menge des verzehrten Olivenöls, sondern auch hinsichtlich zahlloser anderer Variablen: Bruttosozialprodukt, Güte der medizinischen Versorgung, durchschnittliche Familiengröße, durchschnittlicher Alkoholkonsum usw. Alle diese Variablen kommen als alternative Erklärungen für eine höhere Lebenserwartung in bestimmten Ländern in Frage.

Wenn eine Korrelation keinen Beleg für eine Kausalbeziehung darstellt, wie lässt sich dann überhaupt nachweisen, dass eine Variable eine andere Variable ursächlich beeinflusst? Die Antwort haben wir bereits in Kapitel 5 kennen gelernt: Einen solchen Nachweis können Experimente erbringen. Da in Experimenten eine Variable manipuliert und die andere Variable erst anschließend gemessen wird, ist hier nur eine Kausalrichtung möglich. Wenn wir die Variable X manipulieren und die Variable Y anschließend messen, so ist ausgeschlossen, dass die Variable Y die Ursache der Variable X war. Zudem wird durch experimentelle Kontrolltechniken (wie Randomisierung oder Konstanthalten) sichergestellt, dass ein Zusammenhang zwischen den Variablen X und Y nicht auf eine dritte Variable zurückgeht.

Ob das Ergebnis einer Studie als Beleg für eine bestimmte Kausalbeziehung zwischen den untersuchten Variablen interpretiert werden kann, hängt also nicht primär von den statistischen Verfahren ab, die bei der Auswertung verwendet wurden, sondern vom Forschungsdesign, mit dem die Daten erhoben wurden. Auch in Experimenten können korrelative Techniken bei der Datenauswertung eingesetzt werden. Sofern das Experiment intern valide (siehe Kapitel 5) ist und eine nennenswerte Korrelation gefunden wird, kann man aus diesem Ergebnis durchaus auf eine Kausalbeziehung zwischen den Variablen schließen. Entscheidend für diese Interpretationsmöglichkeit ist allerdings, dass durch das experimentelle Design die Kausalrichtung zwischen den beteiligten Variablen eindeutig bestimmt werden kann und durch die Kontrolle von Störvariablen andere Erklärungen für den Zusammenhang zwischen den Variablen ausgeschlossen werden können.

7.6 Partialkorrelation

Tatsächlich stellen Experimente die einzige Möglichkeit dar, Hypothesen über eine Ursache-Wirkungs-Beziehung zwischen zwei Variablen zu testen und dabei *alle* Alternativerklärungen für eine Korrelation zwischen den Variablen auszuschließen. Die Vermutung, dass ein Zusammenhang zwischen zwei Merkmalen durch den Einfluss *einer bestimmten* Drittvariablen hervorgerufen sein könnte, kann aber auch außerhalb von experimentellen Designs mit Hilfe der Technik der Partialkorrelation überprüft werden.

Nehmen wir an, dass in einer deutschen Kleinstadt eine recht starke, positive Korrelation zwischen der Häufigkeit des Kirchgangs und dem Ausmaß der Ausländerfeindlichkeit besteht. Wir vermuten nun aber, dass diese Variablen nicht in einer ursächlichen Beziehung zueinander stehen. Gemäß dieser Vermutung führt also weder ein häufiger Kirchbesuch zu hoher Ausländerfeindlichkeit, noch bewirkt hohe Ausländerfeindlichkeit häufige Kirchgänge. Wir nehmen stattdessen an, dass der Zusammenhang durch eine Drittvariable verursacht wird, nämlich durch das Alter. Diese Annahme kann nur dann korrekt sein, wenn das Alter sowohl mit der Häufigkeit des Kirchgangs als auch mit der Ausländerfeindlichkeit korreliert. Unsere Vermutung impliziert, dass ältere Menschen häufiger in die Kirche gehen und dass sie stärker ausländerfeindlich eingestellt sind. Korreliert das Alter hingegen nicht mit einem oder beiden der anderen Merkmale, so ist ausgeschlossen, dass die Drittvariable „Alter“ den Zusammenhang zwischen der Häufigkeit des Kirchgangs und der Ausländerfeindlichkeit verursacht.

Nehmen wir weiter an, dass unsere Vermutung korrekt ist: Das Alter ist tatsächlich für den beobachteten Zusammenhang verantwortlich. Welche Korrelation zwischen dem Kirchgang und der Ausländerfeindlichkeit sollten wir dann in einer Studie finden, in die nur Menschen eines bestimmten Alters (sagen wir: 50-Jährige) einbezogen werden? Wenn dass Alter konstant gehalten wird, kann es offensichtlich keinen Einfluss auf die anderen Variablen ausüben. Es wäre also zu erwarten, dass die Teilnehmer desselben Alters ähnlich häufig zur Kirche gehen. Zudem sollte auch die Ausländerfeindlichkeit bei allen Teilnehmern ähnlich groß ausgeprägt sein. Folglich würden wir in dieser Stichprobe eine deutlich geringere Korrelation zwischen dem Kirchgang und der Ausländerfeindlichkeit finden als in einer Stichprobe, die Teilnehmer aller Altersstufen enthält.

Generell gilt demnach Folgendes: Sinkt der Zusammenhang zwischen zwei Merkmalen, wenn eine Drittvariable konstant gehalten wird, so kommt diese Drittvariable als Ursache für den Zusammenhang in Frage. Ändert sich die Korrelation zwischen den beiden Merkmalen durch das Konstanthalten der Drittvariablen jedoch nicht, so ist ausgeschlossen, dass die Korrelation auf diese Drittvariable zurückgeht.

Nun ist es nicht erforderlich, die Drittvariable tatsächlich konstant zu halten und beispielsweise den Zusammenhang zwischen der Häufigkeit des Kirchgangs und der Ausländerfeindlichkeit ausschließlich bei 50-Jährigen zu untersuchen. Die Partialkorrelation liefert dieselbe Information, auch wenn die Werte der Drittvariablen frei variieren. Wir könnten also in einer Stichprobe ohne Altersbeschränkung bei jedem Teilnehmer die Ausländerfeindlichkeit, die Häufigkeit des Kirchgangs und das Alter erheben. Die Partialkorrelation gibt dann an, wie hoch der Zusammenhang zwischen dem Kirchgang und der Ausländerfeindlichkeit ausgefallen wäre, wenn wir das Alter konstant gehalten hätten. Alternativ könnte man auch formulieren: Die Partialkorrelation gibt an, wie stark die Korrelation zwischen zwei Variablen ohne den Einfluss einer bestimmten Drittvariablen ausgeprägt wäre. Die Formel für die Partialkorrelation lautet:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{1 - r_{xz}^2} \cdot \sqrt{1 - r_{yz}^2}}$$

Berechnen wir die Partialkorrelation in unserem Beispiel. Gehen wir davon aus, dass in unserer Stichprobe zwischen der Häufigkeit des Kirchgangs (X) und der Ausländerfeindlichkeit (Y) ein Zusammenhang von $r_{xy} = 0{,}42$ besteht. Das Alter (Z) korreliert mit der Häufigkeit des Kirchgangs zu $r_{xz} = 0{,}70$ und mit der Ausländerfeindlichkeit zu $r_{yz} = 0{,}44$. Dann ergibt sich folgende Partialkorrelation:

$$r_{xy.z} = \frac{0{,}42 - 0{,}70 \cdot 0{,}44}{\sqrt{1-0{,}70^2} \cdot \sqrt{1-0{,}44^2}} = \frac{0{,}11}{0{,}71 \cdot 0{,}90} = 0{,}17$$

In unserem Fall verringert sich der Zusammenhang zwischen der Häufigkeit des Kirchgangs und der Ausländerfeindlichkeit also beträchtlich, nachdem der Einfluss des Alters „herauspartialisiert" wurde. Das Alter ist demnach eine mögliche verursachende Drittvariable des Zusammenhangs. Allerdings ist dies noch kein Beweis dafür, dass der Zusammenhang zwischen der Häufigkeit des Kirchgangs und der Ausländerfeindlichkeit tatsächlich auf das Alter zurückgeht! Es ist immer noch möglich, dass der Zusammenhang durch eine andere Drittvariable verursacht wird, die wiederum hoch mit dem Alter korreliert (vielleicht der „Erziehungsstil", dem die Teilnehmer ausgesetzt waren). Wäre die Partialkorrelation jedoch ebenso hoch ausgefallen wie die „einfache" Korrelation zwischen dem Kirchgang und der Ausländerfeindlichkeit, so hätten wir ausschließen können, dass das Alter für diesen Zusammenhang verantwortlich ist.

7.7 Andere Zusammenhangsmaße

Bei der Bestimmung der Stärke des Zusammenhangs zwischen zwei Variablen mit Hilfe des Produkt-Moment-Korrelationskoeffizienten wird berücksichtigt, wie weit die Messwerte von ihrem jeweiligen Mittelwert abweichen. Dies ist natürlich nur sinnvoll, wenn die Abstände zwischen Messwerten auch bedeutsame Information enthalten – oder mit anderen Worten: wenn die Variablen zumindest Intervallskalenniveau aufweisen. Nun trifft man in der Psychologie aber auch auf nominal- oder ordinalskalierte Variablen und die Frage nach dem Zusammenhang solcher Variablen kann durchaus interessant sein. Gibt es etwa einen Zusammenhang zwischen dem Beruf und bestimmten psychischen Störungen? Wenn die Stärke des Zusammenhangs zwischen solchen Variablen ermittelt werden soll, können oder müssen andere Zusammenhangsmaße verwendet werden als der Produkt-Moment Korrelationskoeffizient. Wir wollen zum Abschluss des Kapitels zwei dieser Maße kurz vorstellen (eine Einführung in weitere Korrelationsmaße für nominal- und ordinalskalierte Variablen gibt z.B. Bortz, 2005).

7.7.1 Korrelation zweier dichotomer Merkmale – der Phi-Koeffizient

Dichotome Merkmale sind Variablen, die nur zwei Ausprägungen annehmen können. Beispiele wären also etwa das Geschlecht oder alle Variablen, mit denen kodiert wird, ob ein Merkmal bei einer Person vorhanden ist oder nicht (etwa die Zugehörigkeit zu einer Religionsgemeinschaft oder die Teilnahme an einem Raucherentwöhnungstraining jeweils mit den Ausprägungen „ja" und „nein"). Dichotome Variablen sind somit ein Spezialfall von nominalskalierten Variablen: Ihre Werte bringen lediglich zum Ausdruck, ob eine Person einer bestimmten Kategorie oder Gruppe angehört. Eine Rangordnung der Kategorien im Sinne eines „Mehr“ oder „Weniger“ existiert dagegen nicht.

Gehen wir, um die Bestimmung des Zusammenhangs zwischen zwei dichotomen Merkmalen zu erläutern, von folgendem Beispiel aus: Ein Dozent bittet in einer Vorlesung 100 Studierende, eine Denksportaufgabe zu bearbeiten. Es stellt sich heraus, dass 60 Studierende die Aufgabe lösen können. Zusätzlich geben die Studierenden an, ob sie zuvor bereits an einem Methodenseminar teilgenommen haben. Wie könnte nun die Stärke des Zusammenhangs zwischen den Variablen „Besuch eines Methodenseminars“ und „Lösung der Denksportaufgabe“ ermittelt werden?

Überraschenderweise besteht eine Antwort darin, dass man in diesem Fall den Produkt-Moment-Korrelationskoeffizienten anwenden könnte. Für die Berechnung des Korrelationskoeffizienten ist es natürlich zunächst erforderlich, dass die Kategorienzugehörigkeit der Teilnehmer – wie bei jeder Messung auf Nomimalskalenniveau – in Zahlen übersetzt wird. Wir könnten also etwa allen Studierenden, die die Aufgabe gelöst haben, eine 1 und allen Studierenden, die an der Aufgabe gescheitert sind, eine 0 zuordnen. Ebenso könnten wir für die Merkmalsausprägung „keine Teilnahme an einem Methodenseminar“ eine 0 und für die Ausprägung „Teilnahme an einem Methodenseminar“ eine 1 vergeben. Auf diese Weise würden wir für jeden der 100 Teilnehmer ein Messwertpaar erhalten, das natürlich stets nur aus 0en und / oder 1en bestünde. Mit diesen Messwertpaaren könnten wir nun den Produkt-Moment Korrelationskoeffizienten errechnen – genau wie wir dies im Rechenbeispiel 7.1 mit intervallskalierten Daten getan haben. Das Ergebnis würde die Stärke des Zusammenhangs zwischen dem Besuch eines Methodenseminars und der Lösung der Denksportaufgabe korrekt wiedergeben.

Warum ist der Produkt-Moment-Korrelationskoeffizient in diesem Fall auch auf nominalskalierte Daten anwendbar, obwohl er eigentlich erst ab Intervallskalenniveau definiert ist? Das wesentliche Merkmal intervallskalierter Daten besteht darin, dass gleiche Abstände zwischen den Messwerten gleichen Abständen zwischen den Merkmalsausprägungen entsprechen. Dies ist bei dichotomen, nominalskalierten Merkmalen aber erfüllt. Hier gibt es offensichtlich nur einen Abstand zwischen Messwerten – in unserem Beispiel die Differenz zwischen 0 und 1. Diese Differenz bringt stets den gleichen Unterschied zwischen den Ausprägungen eines Merkmals zum Ausdruck – etwa den Unterschied zwischen der Teilnahme und der Nicht-Teilnahme an einem Methodenseminar. Folglich kann der Produkt-Moment-Korrelationskoeffizient hier eingesetzt werden.

Allerdings gibt es bei der Berechnung einer Korrelation „von Hand“ eine Möglichkeit, den Zusammenhang zwischen dichotomen Merkmalen mit erheblich weniger Aufwand zu bestimmen: Der Phi-Koeffizient führt zu demselben Ergebnis. Die Berechnung des Phi-Koeffizienten setzt zunächst voraus, dass die Daten in Form einer so genannten Vierfeldertafel dargestellt werden. Die ▶Tabelle 7.3 zeigt eine solche Vierfeldertafel. Wie ist diese Tafel zu lesen? Die Zahlenwerte innerhalb der Tabelle geben jeweils die Häufigkeit einer bestimmten Merkmalskombination an. In unserer Stichprobe fanden sich z.B. 32 Personen, die ein Methodenseminar besucht und die Denksportaufgabe gelöst haben. Die Zeilen- und Spaltensummen geben dagegen an, wie oft ein Merkmal *insgesamt* auftrat. In unserem Fall haben insgesamt 40 von 100 Studierenden ein Methodenseminar besucht. 60 Studierende konnten die Denksportaufgabe lösen.

Überlegen wir uns kurz, ob sich anhand dieser Kontingenztafel schon vor der Berechnung eines Korrelationsmaßes erkennen lässt, ob zwischen den Variablen ein Zusammenhang besteht: Offensichtlich konnten in der gesamten Stichprobe 60% der Studierenden die Denksportaufgabe lösen. Von den 40 Studierenden, die an einem Methodenseminar teilgenommen haben, haben 32 die Aufgabe gelöst. Unter den Teilnehmern an einem Methodenseminar beträgt der Anteil der „Löser" demnach 80% (32/40 · 100). Betrachten wir hingegen diejenigen Studierenden, die nicht an einem Methodenseminar teilgenommen haben, so ergibt sich ein anderes Bild: Von den 60 „Nicht-Teilnehmern" konnten 28 die Aufgabe lösen. Dies entspricht einem Anteil von lediglich 46,7%. Der Anteil von Studierenden, die die Aufgabe korrekt beantworten, ist also unter Teilnehmern eines Methodenseminars deutlich größer als unter Nicht-Teilnehmern. Demnach besteht ein Zusammenhang zwischen der Teilnahme am Seminar und der Lösung der Denksportaufgabe.

Tabelle 7.3

		Teilnahme an Methodenseminar
Denksportaufgabe gelöst?	nein	
	ja	

Wie würden die Daten aussehen, wenn zwischen diesen Variablen kein Zusammenhang bestünde? Eine entsprechende Situation ist in ▶Tabelle 7.4 dargestellt. Hier ist der Anteil der „Löser" unter den Studierenden, die an einem Methodenseminar teilgenommen haben, ebenso groß wie unter denjenigen Studierenden, die nicht teilgenommen haben. Er beträgt jeweils 60% – genau wie in der gesamten Stichprobe. Die Chance, dass eine bestimmte Person, die Denksportaufgabe löst, ist in diesem Fall also unabhängig von ihrer Seminarteilnahme. Man würde in dieser Situation daher auch davon sprechen, dass zwischen den Variablen Teilnahme an einem Methodenseminar und Lösung einer Denksportaufgabe „Unabhängigkeit" besteht. Generell ist der Zusammenhang zweier dichotomer Variablen umso stärker ausgeprägt, je deutlicher sich der Anteil einer Merkmalsausprägung auf den beiden Ausprägungen der anderen Variablen unterscheidet. In unserem Beispiel bestünde also ein perfekter Zusammenhang, wenn 100% der Seminarteilnehmer und 0% der Nicht-Teilnehmer die Aufgabe lösen würden.

Tabelle 7.4

		Teilnahme an Methodenseminar
Denksportaufgabe gelöst?	nein	
	ja	
	Spaltensummen	

Berechnung des Phi-Koeffizienten

In der ▶Tabelle 7.5 werden den Häufigkeitsfeldern einer Vierfeldertafel allgemeine Bezeichnungen zugeordnet.

Tabelle 7.5

		Variable X
Variable Y	nein	
	ja	
	Spaltensummen	

Diese Bezeichnungen benötigen wir, um die Formel des Phi-Koeffizienten (ϕ) angeben zu können:

$$\phi = \frac{a \cdot d - b \cdot c}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}$$

Berechnen wir den Phi-Koeffizienten für die Daten in Tabelle 7.3, so ergibt sich also:

$$\phi = \frac{32 \cdot 32 - 8 \cdot 28}{\sqrt{40 \cdot 60 \cdot 60 \cdot 40}} = \frac{800}{2400} = 0{,}33$$

Der Phi-Koeffizient kann genau so interpretiert werden, wie der Produkt-Moment-Korrelationskoeffizient – er führt ja zu demselben Ergebnis. Folgen wir Cohens Konventionen, so würden wir die Korrelation zwischen der Teilnahme an einem Methodenseminar und der Lösung der Denksportaufgabe also als „mittleren Zusammenhang" beurteilen.

7.7.2 Korrelation zweier ordinalskalierter Merkmale – Kendalls Tau

Stellen Sie sich vor, dass Sie und eine Freundin von Ihnen sechs Schokoriegel gemäß Ihren jeweiligen Vorlieben sortiert haben. Sie beide haben dabei jeweils für den „Lieblings-Schokoriegel" den Rangplatz 1 vergeben, der schlechteste Schokoriegel wurde auf Rang 6 gesetzt. Nachdem Sie Ihre beiden Rangreihen verglichen und einige Unstimmigkeiten festgestellt haben, fragen Sie sich, ob sich der Grad der Übereinstimmung zwischen Ihren Vorlieben auch in einer Zahl ausdrücken lässt. Dies ist natürlich möglich: Wir können einfach die Korrelation zwischen den beiden Rangreihen berechnen. Dabei ist allerdings zu beachten, dass es wir es hier mit ordinalskalierten Daten zu tun haben – gleiche Abstände zwischen verschiedenen Rangplätzen bringen nicht unbedingt gleiche Unterschiede in der Vorliebe für verschiedene Schokoriegel zum Ausdruck.

Ein Zusammenhangsmaß, das bei ordinalskalierten Daten häufig angewendet wird, ist Kendalls Tau (τ).[3] Die Berechnung dieses Maßes wird im Rechenbeispiel 7.3 an den Daten aus unserem Schokoriegel-Beispiel illustriert.

Rechenbeispiel 7.3

Vorlieben für Schokoriegel Die ▶Tabelle 7.6 zeigt die Rangplätze, die zwei Urteiler verschiedenen Schokoriegeln aufgrund ihrer Präferenz für diese Schokoriegel zugeordnet haben.

Schokoriegel	A	B	C	D	E	F
Urteiler 1 (X-Variable)	2	5	1	3	4	6
Urteiler 2 (Y-Variable)	1	4	2	5	3	6

Tabelle 7.6: Rangplätze, die zwei Urteiler sechs verschiedenen Schokoriegeln zugeordnet haben.

Der erste Schritt zur Berechnung von Kendalls Tau besteht nun darin, die Schokoriegel gemäß den Rängen, die Urteiler 1 vergeben hat (also gemäß der X-Variablen), anzuordnen. Der Schokoriegel, den Urteiler 1 am meisten bevorzugt, rückt damit auf die Position ganz links. Der von ihm am wenigsten präferierte Schokoriegel nimmt die Position ganz rechts ein. Das Ergebnis dieser „Umsortierung" ist die ▶Tabelle 7.7.

Schokoriegel	C	A	D	E	B	F
Urteiler 1 (X-Variable)	1	2	3	4	5	6
Urteiler 2 (Y-Variable)	2	1	5	3	4	6

Tabelle 7.7: Die Daten aus Tabelle 7.6 sortiert nach der Rangfolge von Urteiler 1. ▶

3 Ein anderes weit verbreitetes Maß für den Zusammenhang zwischen zwei ordinalskalierten Variablen ist Spearmans Rho. Eine Erläuterung zu diesem Maß findet man z.B. bei Bortz (2005).

▸Fortsetzung

Diese umgestellte Tabelle erleichtert es uns, den Grad der Übereinstimmung zwischen Urteiler 1 und Urteiler 2 zu ermitteln. Bei einem perfekten Zusammenhang sollten beide Rangordnungen natürlich vollständig übereinstimmen. Ein nicht perfekter, positiver Zusammenhang zeigt sich darin, dass in der Tabelle 7.7 die niedrigen Ränge auch beim Urteiler 2 eher links stehen, während die hohen Ränge rechts auftreten.

Die Grundlage für die Berechnung von Kendalls Tau bilden alle möglichen Paare von Schokoriegeln. Ermittelt wird die Anzahl der Paare, in denen der Urteiler 2 demselben Schokoriegel den niedrigeren Rangplatz zuweist wie der Urteiler 1 (also den gleichen Schokoriegel bevorzugt). Beginnen wir mit dem Schokoriegel C – demjenigen Objekt, das in der Tabelle 7.7 ganz links steht. Das erste Paar besteht aus den Schokoriegeln C und A. In diesem Paar ist die Rangfolge von Urteiler 2 gegenüber derjenigen von Urteiler 1 offensichtlich vertauscht – der niedrigere Rang (hier 1) taucht beim Urteiler 2 rechts vom höheren Rang (hier 2) auf. Solche Vertauschungen werden als „Inversionen" bezeichnet. Ihnen wird ein Wert von -1 zugeordnet. Bei allen übrigen Paaren mit dem Schokoriegel C (C-D; C-E; C-B; C-F) stimmt die Rangfolge von Urteiler 2 hingegen mit derjenigen von Urteiler 1 überein – der höhere Rang steht in allen diesen Paaren in Tabelle 7.7 rechts. Derartige Übereinstimmungen heißen „Proversionen".

Insgesamt finden wir also bei den Paaren mit dem Schokoriegel C vier Proversionen und eine Inversion. Die Kombinationen von Rängen in diesen Paaren lauten:
Inversionen: (2;1)
Proversionen: (2;5) (2;3) (2;4) (2;6)

Nach demselben Muster werden nun die Proversionen und Inversionen für die übrigen Paare bestimmt. Im nächsten Schritt werden die verbliebenen Paare mit dem Schokoriegel A betrachtet (also die Paare von A mit einem Schokoriegel, der in der Tabelle 7.7. weiter rechts steht). Unter diesen Paaren finden wir ausschließlich Proversionen. Die Kombinationen von Rängen in diesen Paaren lauten:
Proversionen: (1;5) (1;3) (1;4) (1;6)

Unter den verbliebenen Paaren mit dem Schokoriegel D befinden sich zwei Inversionen und eine Proversion:
Inversionen: (5;3) (5;4)
Proversionen: (5;6)

Für die restlichen Paare mit dem Schokoriegel E ergibt sich das folgende Muster:
Proversionen: (3;4) (3;6)

Schließlich tritt beim letzten Paar B-F ebenfalls eine Proversion auf:
(4;6) ▸

▸Fortsetzung

Unter allen Paaren von Schokoriegeln finden wir demnach insgesamt 12 Proversionen und 3 Inversionen – bei 12 Paaren bevorzugen beide Urteiler den gleichen Schokoriegel, bei 3 Paaren gehen ihre Vorleiben dagegen auseinander. Bei der Bestimmung von Kendalls Tau wird nun die Differenz zwischen der Anzahl der Proversionen (P) und der Anzahl der Inversionen (I) berechnet. Diese Differenz wird als S bezeichnet:

$$S = P - I$$

In unserem Fall ergibt sich $S = 12 - 3 = 9$. Diese Differenz wird relativiert an der maximal möglichen Anzahl an Proversionen. Die maximale Anzahl an Proversionen würde natürlich dann erreicht werden, wenn die Urteiler bei *allen* Paaren übereinstimmen würden. Die Anzahl aller Paare ergibt sich nach der Formel

$$n\,(n-1)\,/\,2$$

wobei mit n die Anzahl der Objekte bezeichnet wird. Die Anzahl aller Paare beläuft sich in unserem Fall demnach auf $6 \cdot (6-1)\,/\,2 = 15$.
Die Formel für Kendalls Tau (τ) lautet entsprechend:

$$\tau = \frac{S}{n \cdot (n-1)/2}$$

Der Zusammenhang zwischen den Vorlieben der beiden Urteiler beträgt also:

$$\tau = \frac{9}{6 \cdot (6-1)/2} = \frac{9}{15} = 0{,}60$$

Kendalls Tau kann, genau wie der Produkt-Moment-Korrelationskoeffizient, Werte zwischen −1 und 1 annehmen. Sofern die beiden Rangreihen exakt übereinstimmen, besteht zwischen den Variablen natürlich ein perfekter, positiver Zusammenhang. In diesem Fall erreicht auch die Anzahl der Proversionen den maximal möglichen Wert und Kendalls Tau beträgt 1. Kendalls Tau wird umso niedriger ausfallen, je größer die Anzahl der Inversionen ist. Kommen gleich viele Proversionen und Inversionen vor, so beträgt Kendalls Tau 0. Der Wert −1 wird schließlich erreicht, wenn ausschließlich Inversionen auftreten. Dies ist dann der Fall, wenn die beiden Rangreihen „umgekehrt" verlaufen: Dasselbe Objekt erhält also auf der Variable X den besten Rangplatz und auf der Variable Y den schlechtesten; einem anderen Objekt wird auf der Variable X der zweitbeste und auf der Variable Y der zweitschlechteste Rangplatz zugewiesen usw.

Die Berechnung von Kendalls Tau kann natürlich auch dann sinnvoll sein, wenn die Daten ursprünglich nicht in Form von Rangreihen vorliegen. Betrachten wir ein Beispiel: Nehmen wir an, dass wir untersuchen wollen, wie stark die Notenurteile zweier Lehrer bei 10 verschiedenen Klausuren zusammenhängen. Nun ist die Auffassung, dass Noten kein Intervallskalenniveau aufweisen, weit verbreitet und recht gut begründet.

Sofern wir uns dieser Auffassung anschließen, sollten wir auf die Berechnung des Produkt-Moment Korrelationskoeffizienten verzichten. Stattdessen könnten wir eine „Rangkorrelation“ zwischen den Notenurteilen bestimmen – also beispielsweise Kendalls Tau. Bei der Berechnung dieses Maßes könnten wir – aus Gründen der Übersichtlichkeit – zunächst die ursprünglichen Notenwerte in Ränge umwandeln. War also einer der Lehrer der Ansicht, dass die beiden besten Klausuren mit den Noten „2−„ und „3+“ zu bewerten sind, so würden wir diesen Klausuren die Rangplätze 1 und 2 zuordnen.[4] Letztlich ist für die Berechnung von Kendalls Tau aber nur die korrekte Bestimmung von Proversionen und Inversionen relevant. Diese kann natürlich auch anhand der ursprünglichen Notenwerte vorgenommen werden.

ZUSAMMENFASSUNG

In diesem Kapitel werden Verfahren erörtert, mit denen der Zusammenhang – bzw. die *Korrelation* – zwischen zwei Variablen analysiert werden kann. Eine Möglichkeit, Zusammenhänge zwischen Variablen grafisch darzustellen, bieten Streudiagramme. In solchen Diagrammen wird eine Variable auf der x-Achse und die andere Variable auf der y-Achse eines Koordinatensystems abgetragen. Die Messwerte einer Person auf beiden Variablen werden als ein Punkt in das Koordinatensystem eingezeichnet.

Mit Hilfe von Streudiagrammen lässt sich zunächst die Form und Richtung von Zusammenhängen bestimmen. Folgen die Punkte in etwa einer Geraden, so besteht zwischen den Variablen ein linearer Zusammenhang.

Beschreibt die „Punktwolke“ eher die Form einer Kurve, so handelt es sich um einen kurvilinearen Zusammenhang. Wenn die Punkte kreisförmig angeordnet sind, so dass sich keinerlei Trend in den Daten erkennen lässt, sind die Variablen unkorreliert. Bei linearen Korrelationen wird weiter zwischen einer positiven und negativen Richtung unterschieden. Gehen hohe Werte auf einer Variablen mit hohen Werten auf der anderen Variablen einher, so wird dies als positive Korrelation bezeichnet. Von einer negativen Korrelation spricht man hingegen, wenn hohe Werte auf einer Variablen mit niedrigen Werten auf der anderen Variablen verbunden sind.

Der Produkt-Moment-Korrelationskoeffizient (r) ist ein Maß für die Stärke und Richtung des linearen Zusammenhangs zwischen zwei intervallskalierten Variablen. Der Korrelationskoeffizient nimmt Werte zwischen -1 und $+1$ an. Bei perfekten, positiven Zusammenhängen beträgt der Korrelationskoeffizient $+1$, bei perfekten negativen Zusammenhängen -1. Der Wert $r = 0$ signalisiert, dass zwischen zwei Variablen kein Zusammenhang besteht. ▶

4 Dabei kann es natürlich vorkommen, dass ein Lehrer die gleiche Note mehrfach vergibt. Klausuren mit identischen Noten wäre auch der gleiche Rang zuzuweisen. Würden beispielsweise zwei Klausuren mit derselben Note die Rangplätze 5 und 6 einnehmen, so würde beiden Klausuren der gemittelte Rangplatz 5,5 zugeordnet. Treten solche Rangbindungen (oder „ties“) auf, so muss bei der Berechnung von Kendalls Tau eine korrigierte Formel verwendet werden. Eine Erläuterung dieser korrigierten Formel geben z.B. Siegel und Castellan (1988).

▶Fortsetzung

Der Produkt-Moment Korrelationskoeffizient kann durch verschiedene Faktoren „verzerrt" werden. Daher ist bei der Interpretation des Koeffizienten darauf zu achten, ob solche Verzerrungen eingetreten sind. Ausreißerwerte können den Korrelationskoeffizienten „künstlich" vergrößern oder verkleinern. Eine Einschränkung der Variabilität einer oder beider Variablen führt dagegen stets zu einer Verminderung des Koeffizienten. Besteht eine Stichprobe aus heterogenen Untergruppen, so kann der Korrelationskoeffizient in der Gesamtstichprobe deutlich von den Koeffizienten in den Untergruppen abweichen. In diesem Fall sollten die Untergruppen gesondert analysiert werden.

Bei der Interpretation des Produkt-Moment-Korrelationskoeffizienten sollte zudem immer beachtet werden, dass ein Zusammenhang zwischen zwei Variablen keine eindeutige Aussage über eine Ursache-Wirkungs-Beziehung zwischen den Variablen erlaubt. Eine Korrelation zwischen Variablen ist lediglich eine notwendige, aber keine hinreichende Voraussetzung für die Schlussfolgerung, dass die Variable X die Variable Y ursächlich beeinflusst.

Eine Partialkorrelation gibt an, wie stark der Zusammenhang zwischen zwei Merkmalen ausgeprägt wäre, wenn eine bestimmte Drittvariable keinen Einfluss auf die Merkmale hätte. Ist die Partialkorrelation ebenso hoch wie die Korrelation zwischen den Merkmalen, so kann ausgeschlossen werden, dass der Zusammenhang durch die „herauspartialisierte" Drittvariable verursacht wird.

Soll der Zusammenhang zwischen zwei Variablen, die nicht mindestens Intervallskalenniveau aufweisen, bestimmt werden, so können oder müssen andere Maße eingesetzt werden als der Korrelationskoeffizient. Ein Maß für den Zusammenhang zwischen zwei dichotomen, nominalskalierten Variablen ist der Phi-Koeffizient. Bei ordinalskalierten Variablen wird oftmals Kendalls Tau angewendet.

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

Weiterführende Literatur

Cohen, J., Cohen, P. & West, S.G. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Umfassende, wenig „mathematische" Darstellung der Korrelations- und Regressionsrechnung (siehe auch Kapitel 8) und ihrer Anwendungsmöglichkeiten in der Psychologie.

Übungsaufgaben mit Lösungen sowie weitere Informationen zu diesem Buchkapitel finden Sie auf der Companion Website zum Buch unter *http://www.pearson-studium.de*