

# Vorwort

Einzelstatistik hatte vor noch 25 bis 30 Jahren großes Interesse in den Sozialwissenschaften gefunden – Beleg dafür sind etwa im angloamerikanischen Raum die bekannten Monographien von Glass, Willson u. Gottman (1975), Gottman (1981), Barlow u. Hersen (1984) oder das deutschsprachige viel beachtete Sammelwerk von Petermann u. Hehl (1979); mittlerweile ist Einzelfallanalyse als methodisches Werkzeug offenbar in den Hintergrund getreten. Dabei sind die wissenschaftlichen Fragestellungen und Datensätze, die einzelstatistische Herangehensweise erfordern, nach wie vor aktuell, im zunehmenden Bemühen um eine empirisch fundierte Psychotherapieforschung sogar vielleicht von größerer Bedeutung denn je: Nachweis von Effekten an behandelten Personen, die sich nicht zu größeren Interventionsgruppen zusammenfassen lassen, sondern die als alleinige Untersuchungsobjekte oder bestenfalls als homogene Kleinstgruppen angesehen werden müssen.

Dass die einzelstatistische Forschung (wenigstens in den Sozialwissenschaften) nicht die eigentlich zu erwartende Weiterentwicklung genommen hat, liegt sicher wesentlich daran, dass die statistische Auswertung im Einzelfall mit gewissen methodischen Schwierigkeiten behaftet ist – hier ist in erster Linie die serielle Abhängigkeit der Daten zu nennen – und ihr Verständnis nicht leicht fällt. Insbesondere gilt: ForscherInnen, deren Daten vor allem eine solche Herangehensweise erfordern, nämlich häufig psychotherapeutisch oder beraterisch Tätige, sind nicht unbedingt identisch mit jenen, welche sich in die Feinheiten der Einzelstatistik einzuarbeiten bereit sind.

Auf diesem Hintergrund ist die Idee zu dem vorliegenden kleinen Buch gewachsen: Um einzelstatistische Auswertung ist bei vielen Datensätzen nicht herzukommen – es sei denn, man verschenkt Information oder riskiert elementare methodische Kritik; andererseits muss eine solche Auswertung durchzuführen sein ohne ungebührlich lange Einarbeitung in subtile Zeitreihenmodelle. Hier wird deshalb versucht, eine knappe, hoffentlich verständliche, teilweise bewusst simplifizierende Einführung in die Einzelstatistik zu leisten. Insbesondere wurde weitgehend auf die eleganten, aber für Anfänger befremdenden Darstellungen mittels Matrizen und Operatoren verzichtet, dies bestenfalls in den Anmerkungen angedeutet. In die Anmerkungen wurde auch ein Großteil der sonstigen mathematischen Begründungen verfrachtet. Diese zu lesen, wird natürlich herzlich aufgefordert; den Gang der Argumentation sollte man aber prinzipiell auch ohne sie verstehen. Explizit sei zugestanden, dass die Ausführungen an einigen Stellen nicht „exakt“ sind: So wurde nicht immer streng zwischen einer empirischen Zeitreihe und einer theoretischen (einem „Zeitreihenprozess“) unterschieden, nicht konsequent Schätzwerte mit einem „Dachsymbol“ kenntlich gemacht, Schätzwerte als Identitäten aufgeführt und zuweilen Voraussetzungen etwas lasch gehabt. Dafür dürfte diese Darstellung aber

für statistisch wenig vorgebildete Personen leichter zu verstehen sein, was mir einige Nachlässigkeiten zu rechtfertigen scheint.

Großen Wert wurde auf Rechenbeispiele gelegt: Jeder sollte beispielsweise selbst einmal die Erstellung eines zeitreihenanalytischen Modells aus den Kennwerten des konkreten individuellen Datensatzes nachvollziehen können und die Residuen bestimmen, wenn die im Modell implizierten Zusammenhänge zwischen den Einzeldaten eliminiert werden. Diese Rechenbeispiele bestehen aus kurzen, fiktiven Datenreihen, und ihre Auswertung ist des mittels einfacher Taschenrechner per Hand möglich. Dass solche Beispiele kritisierbar sind, beispielsweise sich oft gewisse Voraussetzungen nicht überprüfen lassen (etwa die Stationarität), sei gerne konzediert.

Von den umfangreichen Anmerkungen abgesehen, ist der eigentliche Text vergleichsweise kurz. Bedenkt man noch, dass zudem eine kurze Einführung in einzelfallanalytische Versuchspläne und Messinstrumente gegeben wird, so bleibt für das, was den Inhalt umfangreicher Bücher ausmacht, nämlich die Zeitreihenanalyse, wenig Platz. Eingehende Ausführungen zu diesem Thema dürfen also nicht erwartet werden; für Vertiefung wird im Text auf weiter gehende Literatur verwiesen, nicht zuletzt auf die bedauerlicherweise nie mehr aktualisierte Monographie von Gottman (1981), der prinzipiell diese Darstellung immer wieder auf längere Abschnitte folgt.

Dank eines glücklichen Zufalls konnte ich von der Hilfe R. Schlittgens profitieren, der selbst mehrere, in zahlreichen Auflagen erschienene Bücher über Zeitreihenanalyse verfasst hat und mir bei verschiedenen Problemen wesentliche Lösungsansätze unterbreitete. Dem Beltz Verlag, insbesondere Frau H. Berger, Frau M. Raddecki und Frau A. Renz danke ich sehr für das Publikationsangebot sowie die fachmännische Hilfe bei Abfassung und Texterstellung. Wie so oft, wäre auch hier ohne das Engagement und die profunden Computerkenntnisse von I. Böschen und H. Singmann die Texterstellung gar nicht möglich gewesen; wie schon früher, ist mir auch mein Kollege Reinhold Schwab diesbezüglich von großer Hilfe gewesen. Meine liebe Frau Carmen, welche die Abfassung vieler Bücher zuvor ertragen musste, hat nachweislich bei diesem am meisten gelitten; ich kann nur hoffen, dass ihr Verständnis trotzdem noch für weitere Bücher von mir reicht.

Hamburg, im November 2007

Thomas Köhler

# 1 Charakteristik und Indikation von Einzelfallanalysen

## 1.1 Definitionen

Eine *einzelfallanalytische Untersuchung* (Studie) oder *Einzelfallanalyse* wird dann durchgeführt, wenn explizit die *Daten eines Individuums* (allgemeiner: einer Untersuchungseinheit<sup>1</sup>) in einen *Zusammenhang gebracht werden*. In einfachster Form ist dies bereits der Fall, wenn beispielsweise mittels eines Fragebogens der Intelligenzwert einer Person erhoben wird, um eine Entscheidung über die Berufsausbildung zu treffen; ebenfalls eine solche einfache Einzelstudie stellt die laborchemische Bestimmung eines Blutwerts bei einem Patienten dar, um daraufhin gegebenenfalls eine Therapie einzuleiten. Nicht diese elementaren Einzelfallstudien<sup>2</sup> sollen im Weiteren betrachtet werden, ebenso wenig jene, in denen ein Individuum mit verschiedenen Instrumenten zu einem Zeitpunkt (in einer Situation) zur Erstellung eines Einzelprofils untersucht wird. Uns beschäftigen hier jene Studien, bei denen mit mindestens einem Instrument zu *verschiedenen Zeitpunkten* (in *verschiedenen Situationen*) die Untersuchung des betreffenden Individuums erfolgt und diese Daten in einen inneren (intraindividuellen) Zusammenhang gebracht werden.

Werden die untersuchten Variablen nicht explizit mit Hilfe von Zahlen quantifiziert, wird also beispielsweise nur von großen Trennungsängsten oder Bindungsunfähigkeit gesprochen, handelt es sich um eine *qualitative* Einzelfallstudie, was die sicher am meisten verbreitete Form von Einzelfalluntersuchung im oben definierten Sinne darstellt. Bei *quantitativen* Einzelfallstudien werden die Ausprägungen in den interessierenden Variablen skaliert, d. h. mit Zahlen versehen, und diese Zahlenwerte in Verbindung gesetzt – dabei sei angemerkt, dass auch bei vorwiegend qualitativen Einzelfallanalysen häufig diverse quantitative Daten erhoben werden, sofern dafür geeignete Messinstrumente vorliegen, diese aber in qualitativen Begriffen (wie große soziale Ängstlichkeit, geringes Schuldbewusstsein) in das Untersuchungsergebnis eingehen. Eine *statistische Einzelfalluntersuchung* ist eine Studie *quantitativer* Art, bei der die erhaltenen Daten des betrachteten Individuums mittels *statistischer Methoden* intraindividuell (innerhalb der Person) zueinander in Verbindung gesetzt werden; dies kann rein deskriptiv geschehen, indem etwa Mittelwerte in der Variable „soziale Angst“ für Zeitpunkte vor und nach Therapie angegeben werden – wohlgemerkt, nicht durch *Mittelung* über verschiedene Personen, sondern *über verschiedene Erhebungszeitpunkte bei der einzigen als Untersuchungsobjekt dienenden Person*. Über diese Deskription hinaus kann versucht werden, die Daten auf Signifikanz zu testen, also bei dieser Person ein zufälliges Zustandekommen von Mittelwertunterschieden vor und nach Therapie auszuschließen – es sei hier schon betont, dass die üblichen gruppenstatistischen Testverfahren (etwa *t*-Test, Va-

riananalyse) dabei nicht oder nur unter starken Einschränkungen bzw. nach zusätzlichen Analysen eingesetzt werden können.

Das Gegenstück zur *Einzelfallstatistik* ist die *Gruppen-* oder *Aggregatstatistik*. Beide Begriffe sind wenig geläufig; Aggregatstatistik wird nämlich (etwas ungenau) mit Statistik schlechthin gleichgesetzt. Mittels aggregat- oder gruppenstatistischer Verfahren versucht man, Aussagen über ein *Aggregat*, d. h. eine *Menge von Untersuchungsobjekten in ihrer Gesamtheit*, zu treffen. Ein Beispiel wäre die Bildung des Mittelwerts oder Bestimmung der Varianz einer Variable in einer Stichprobe; die Werte der einzelnen Personen gehen zwar in die Berechnung ein, tauchen aber nicht mehr im Rechenresultat auf; statt dessen ergeben sich *Kennwerte* für das *Aggregat* in seiner Gesamtheit (Stichprobenkennwerte). Unter bestimmten Bedingungen können bekanntlich diese Werte von Stichproben auf Populationen übertragen werden (Ziel der *schließenden* oder *Inferenzstatistik*), z. B. mit einem zu konzedierenden Fehler („Vertrauensbereich“) der Stichprobenmittelwert als Schätzung des Populationsmittelwertes dienen. Dies ist der große Vorteil der sonst oft inhaltlich wenig befriedigenden Aggregataussagen, nämlich dass sie von einem überschaubaren *Subaggregat*, der *Stichprobe*, auf das der Untersuchung nicht zugängliche *Gesamtaggregat*, die *Population*, übertragen werden können (*statistischer Induktionsschluss*).

Bei einer *einzelfallanalytischen Untersuchung (Einzelfallstudie)* wird versucht, an einer Untersuchungseinheit (im häufigen Fall: einem Individuum) erhobene Daten in einen (*intraindividuellen*) *Zusammenhang* zu bringen. Eine *quantitative Einzelfallstudie* betrachtet *skalierte*, d. h. mit Zahlen versehene *Daten*. Im Rahmen *statistischer Einzelfallstudien* werden Verfahren der Statistik benutzt, um *intraindividuelle Zusammenhänge* zu beschreiben oder/und *auf Zufälligkeit* zu überprüfen; es sei schon darauf hingewiesen, dass die üblichen *Verfahren der Gruppenstatistik* dafür nur *mit Einschränkungen* geeignet sind.

## 1.2 Arten wissenschaftlicher Hypothesen und ihre Überprüfung

Lehrreich ist die Überlegung, welche Arten von wissenschaftlichen Hypothesen mit welchen der beiden genannten statistischen Vorgehensweisen (Einzelfall- oder Aggregatstatistik) überprüft werden können; man kommt nämlich zum überraschenden Resultat, dass bei vielen der in Psychologie, Pädagogik oder auch Medizin gängigen Hypothesen eine *einzelfallstatistische, nicht eine gruppenstatistische Herangehensweise angebracht* ist. Bunge (1967, S. 238, zitiert nach Westmeyer, 1979) unterscheidet – hier vereinfacht wiedergegeben – vier große Typen wissenschaftlicher Aussagen (bzw. Hypothesen, solange der Wahrheitsgehalt noch nicht fest steht):

- Eine *singuläre* Aussage bezieht sich auf ein einziges Objekt (im hier vornehmlich interessierenden Fall: auf ein bestimmtes Individuum), etwa: Patient A leidet an myeloischer Leukämie; oder: Person B hat nach der Therapie signifikant häufiger pro Tag soziale Kontakte als vor der Behandlung<sup>3</sup>.
- Eine *Existenzhypothese* behauptet die Existenz eines oder mehrerer Objekte, für welche eine bestimmte singuläre Aussage zutrifft; beispielsweise: Es gibt Perso-

nen, bei denen sich im Anschluss an eine Wallfahrt ein zuvor gesicherter Tumor nicht mehr nachweisen ließ; oder: Bestimmte Personen sind in der Lage, durch Meditation ihre Pulsfrequenz um 30% zu senken.

- Eine *universelle* Hypothese ist eine All-Aussage, z. B. Freuds berühmte Verführungstheorie: „Ich stelle also die Behauptung auf, zugrunde jedes Falles von Hysterie befinden sich [...] **ein oder mehrere Erlebnisse von vorzeitiger sexueller Erfahrung**, die der frühesten Jugend angehören.“ Oder: Bei allen Personen mit Down-Syndrom lässt sich eine Trisomie 21 nachweisen. Von den universellen Aussagen sind die *Aggregataussagen* klar *zu unterscheiden*.
- *Aggregataussagen* machen – wie oben erläutert – eine Aussage über ein *Aggregat von Objekten* (typischerweise Personen) in *seiner Gesamtheit*, nicht über die *Einzelobjekte*. Etwa: Die mittlere Lebenserwartung der Neugeborenen in Deutschland beträgt 79,6 Jahre; oder: Die Rückfallwahrscheinlichkeit bei Sexualstraftätern in einer bestimmten Region beträgt 45%; oder: 72% der Lehrer werden vorzeitig pensioniert; oder: Die durchschnittliche deutsche Familie hat 1,3 Kinder; oder: Die Standardabweichung des Alters in der untersuchten Stichprobe betrug 4,34 Jahre. Es leuchtet unmittelbar ein, dass eine solche Aggregataussage nicht auf die Individuen des Aggregats übertragen werden darf: So beträgt die Rückfallwahrscheinlichkeit für einen bestimmten Sexualstraftäter entweder 100% oder 0%, d. h. er wird entweder rückfällig oder nicht; ein bestimmtes Individuum wird eine Lebenserwartung von ziemlich sicher nicht genau 79,6 Jahren haben; keine einzige deutsche Familie hat 1,3 Kinder; die Standardabweichung des Alters in einer Stichprobe ist überhaupt für ein einzelnes Stichprobenelement nicht definiert.

Die Frage, an welche Hypothesentypen aggregatstatistisch oder besser einzelfallanalytisch heranzugehen ist, lässt sich rasch beantworten. Der Wahrheitsgehalt von *singulären*, *Existenz-* und *universellen* Hypothesen kann nur *einzelfallanalytisch* überprüft werden. Die singuläre Aussage, Patient A leidet an myeloischer Leukämie, setzt genaue Untersuchung dieses Individuums voraus, insbesondere Erhebung des Blutbildes, am besten mehrfach, um mögliche Irrtümer auszuschließen. An einem Aggregat erhobene Daten werden dabei lediglich insofern gebraucht, als sie die für die Beurteilung wertvollen *Referenzwerte* liefern; die dazu notwendige Untersuchung des Aggregats hat aber typischerweise lange vorher und unabhängig vom Fall des betrachteten Patienten A stattgefunden, diente zur Ermittlung von Normwerten für Blutkörperchen in der Population. Weniger einfach gestaltet sich die Überprüfung der Hypothese bezüglich Person B: Hier ist in jedem Fall zumindest eine zweifache Erhebung der sozialen Kontakte erforderlich, nämlich vor und nach der Therapie. Bekanntlich schwanken diese beträchtlich über die Erhebungszeitpunkte (abhängig u. a. von Jahreszeit, Wetter, Arbeitsbelastung), sodass Vergleich zweier Werte so gut wie gar keine Aussagekraft besitzt; erst die mehrfache Erhebung der Zahl sozialer Kontakte vor und nach Therapie (am besten über längere Intervalle, in denen alle denkbaren Varianten von Jahreszeit, Wetter und Arbeitsbelastungen in jedem der verglichenen Zeiträume vorkommen) sowie eine sorgsame einzelfallstatis-

tische Auswertung, die insbesondere die serielle Abhängigkeit der Daten in Rechnung setzen muss, können den schlüssigen Nachweis erbringen, dass sich nach Therapie die Anzahl sozialer Kontakte in überzufälligem Ausmaß erhöht hat.

Dass die Überprüfung einer Existenzhypothese, die ja eine Variante einer singulären darstellt, ebenfalls einzelfallanalytisch zu geschehen hat, ist trivial. So müssten eben bei einer Person, deren Tumor sich angeblich nach einer Wallfahrt zurückgebildet hat, Befunde vorher und nachher in Beziehung gesetzt werden.

Eine *universelle Aussage* stellt eine *Menge von singulären Aussagen* dar, die für *alle* Mitglieder einer Gruppe (Stichprobe oder Population) gelten. Die Universalität einer Aussage für Elemente einer nicht in ihrer Gesamtheit der Untersuchung zugänglichen Menge zu beweisen, stellt bekanntlich ein erhebliches logisches Problem dar; will man gleichwohl den Versuch unternehmen, muss dies natürlich durch *einzelfallanalytische Studien* geschehen, ebenso wie Widerlegung einer universellen Aussage nur durch Nachweis eines nicht hypothesenkonformen Einzelfalls gelingt.

*Aggregataussagen* sind die einzigen, die gruppen- oder aggregatstatistische Herangehensweise erfordern, wobei die Logik des statistischen Induktionsschlusses es gestattet, sich bei der Untersuchung – anders als bei der Verifizierung einer universellen Hypothese – auf ein *Subaggregat zu beschränken*, eine so genannte *Stichprobe*. Vorausgesetzt, dieses Subaggregat ist repräsentativ für die ins Auge gefasste Population, lässt sich der Stichprobenbefund auf letztere verallgemeinern, wobei man entweder einen bestimmten Schätzfehler konzedieren muss (z. B. bei der Schätzung des Populations- aus dem Stichprobenmittelwert) oder im Falle komparativer Aussagen (etwa bei prä-post-Vergleichen) eine gewisse Irrtumswahrscheinlichkeit der verallgemeinernden Aussage zugibt.

Wissenschaftliche Hypothesen lassen sich in *singuläre, Existenz- und universelle Hypothesen* unterteilen; daneben gibt es die *Aggregathypothesen*, welche Aussagen über eine *Menge in ihrer Gesamtheit*, nicht über deren einzelne Elemente, machen. Überprüfung von Hypothesen der ersten Kategorie kann nur *einzelfallanalytisch* geschehen; lediglich *Aggregathypothesen* müssen mit *gruppenanalytischen Methoden*, beispielsweise gruppenstatistischen Verfahren, angegangen werden.

### 1.3 Aggregierung einzelfallanalytisch gewonnener Aussagen

Einzelfall- und aggregatstatistische Auswertung schließen sich keineswegs aus: Dass zum Zwecke der Verallgemeinerung eine Aggregierung singulärer Aussagen wenn irgendwie möglich versucht werden sollte, ist selbstverständlich, es sei denn, man will sich mit singulären Aussagen des Typs begnügen, bei Patienten A, B und D habe sich eine einzelfallstatistisch überzufällige Verbesserung ergeben, bei Patient C nicht. In vielen Fällen wird die Schärfe des gruppenstatistischen Tests (seine Power) durch einzelfallanalytische Vorbehandlung der Daten erhöht<sup>4</sup>; ein solcher Fall sei anhand fiktiver Daten besprochen.

Bei 15 Personen  $P_i$  wurde vor einer Therapie dreimal in regelmäßigen Abständen mit einem Fragebogen die Intensität der sozialen Kontakte erhoben; das geschah ebenfalls dreimal, nachdem die Probanden ein soziales Kompetenztraining absolviert hatten. Tabelle 1.1 zeigt die erhaltenen Rohwerte (Scores)  $x_{ij}$  und  $y_{ij}$  sowie erste Zwischenwerte der Analyse. In Spalte 4 finden sich die Differenzen der für jeden Probanden über die jeweils drei Messzeitpunkte gemittelten Phasenscores (also  $\bar{y}_i - \bar{x}_i$ ). In der sechsten Spalte steht die Zahl der in der Phase nach Therapie (im Vergleich zum individuellen Mittelwert) erhöhten Messwerte für Intensität sozialer Kontakte (bezeichnet als Variable  $V$  mit den Probandenwerten  $v_i$ ). Proband 1 hat über alle 6 Messzeitpunkte gemittelt eine durchschnittliche Intensität sozialer Kontakte von 6,5; alle post-Therapie-Werte liegen darüber, also ergibt sich  $v_1 = 3$ . (In den fiktiven Daten wurde der Fall ausgeschlossen, dass Rohwerte  $x_{ij}$  und  $y_{ij}$  genau mit dem individuellen Durchschnitt identisch sind, weiter, dass mehr als 3 von den 6 Werten überdurchschnittlich hoch liegen.)

**Tabelle 1.1: Scores für soziale Kontakte vor und nach Therapie (fiktive Daten)**

Nummer des Pb $i$ 1,2,...,15	Scores vor Therapie $x_{ij}$ zu Messzeitpunkten $j = 1,2,3$	Scores nach Therapie $y_{ij}$ zu Messzeitpunkten $j = 1,2,3$	Differenz zwischen durchschnittlichem post- und prä-Score $z_i = \frac{1}{3} \sum_{j=1}^3 y_{ij} - \frac{1}{3} \sum_{j=1}^3 x_{ij}$	Durchschnittlicher Wert über alle Messungen beider Phasen $z_i = \frac{1}{6} \left( \sum_{j=1}^3 y_{ij} + \sum_{j=1}^3 x_{ij} \right)$	Anzahl erhöhter Werte in post-Therapie-Phase ( $v_i$ )
1	5; 5; 5	7; 8; 9	3	6,5	3
2	7; 6; 5	2; 5; 8	-1	5,5	1
3	7; 7; 7	2; 3; 4	-4	5	0
4	3; 5; 7	4; 8; 9	2	6	2
5	4; 4; 4	8; 8; 8	4	6	3
6	8; 8; 8	4; 4; 4	-4	6	0
7	2; 5; 8	3; 9; 9	2	6	2
8	1; 1; 1	7; 7; 7	6	4	3
9	6; 6; 6	4; 4; 4	-2	5	0
10	8; 4; 9	7; 3; 2	-3	5,5	1
11	6; 6; 6	5; 5; 5	-1	5,5	0
12	5; 5; 5	6; 6; 6	1	5,5	3
13	3; 3; 3	5; 5; 5	2	4	3
14	5; 8; 8	3; 7; 2	-3	5,5	1
15	2; 8; 5	7; 6; 5	1	5,5	2
	$\bar{x} = 5,33$	$\bar{y} = 5,53$	$\bar{d} = 0,2$		

Bei der üblichen aggregatstatistischen Auswertung würden wahrscheinlich zunächst die Werte der einzelnen Probanden in den beiden Phasen des Versuchs gemittelt – dieses Vorgehen liefert zwar einen weniger fehlerbehafteten Wert als die Einzelmessungen; jedoch wird hierbei die Information über die intraindividuelle Varianz der Scores für Intensität sozialer Kontakte verschenkt. Sodann würde üblicherweise die Differenz der durchschnittlichen Werte zwischen den Phasen vor und nach Therapie berechnet und mittels des  $t$ -Tests für korrelierende Stichproben diese Unterschiede auf Signifikanz getestet, also überprüft, ob die Differenzen  $d_i$  von einer Normalteilung (bzw.  $t$ -Verteilung) um 0 abweichen. Da die Differenzen aber offensichtlich nicht normalverteilt sind, dürfte man alternativ hier auf den Wilcoxon-Test als non-parametrisches Verfahren ausweichen. Dabei ergibt sich für den kleineren der beiden T-Werte des Tests 57,5, was natürlich weit von jeglicher Signifikanz entfernt ist. Konventionelle aggregatstatistische Auswertung liefert also keinen Anhalt dafür, dass das Kompetenztraining die Intensität sozialer Kontakte verändert hat.

Anders ist die Situation, wenn die Daten einzelfallanalytisch aufbereitet werden, also die bei der Mittelung verloren gehende intraindividuelle Variabilität berücksichtigt wird. Zwar lässt sich an den wenigen vorliegenden Daten keine individuelle Signifikanzprüfung durchführen<sup>5</sup>; immerhin wurde aber für jeden Probanden ein Kennwert erhalten, welcher die Veränderungen seiner Sozialkontakte in den beiden Phasen charakterisiert. Bei Proband 8 liegt offensichtlich ein recht konsistenter Anstieg der sozialen Kontaktfreudigkeit nach Therapie vor, bei Proband 9 ist genau das Gegenteil zu beobachten. Probanden 2 und 15 haben keine nennenswerten Veränderungen in die eine oder andere Richtung gezeigt.

Unter Zufallsbedingungen (fehlendem Einfluss der Therapie auf die Intensität sozialer Kontakte) sollte die Wahrscheinlichkeit für einen überdurchschnittlichen Wert – nennen wir solche Werte  $b$ , unterdurchschnittliche Werte  $a$  – in jeder Messung den Wert 0,5 annehmen, unabhängig ob diese vor oder nach Therapie durchgeführt wurde. Die Wahrscheinlichkeit für drei erhöhte Werte nach Therapie (also für einen Wert 3 in Spalte 6) würde dann  $0,5^3$  betragen, für 0 erhöhte Scores nach Therapie (entsprechend einem Wert von 0 in Spalte 6) gleichfalls  $0,5^3$ . Den Wert 2 erhält man bei den Konstellationen  $bba$  oder  $bab$  oder  $abb$  (also erster und zweiter Scores der post-Therapie-Phase erhöht, oder erster und dritter, usw.); die Zufallswahrscheinlichkeit dafür ergibt sich als  $3/8$ . Ein Wert von 1 in Spalte 6 ergibt sich, wenn nur eine der drei Messungen nach Therapie erhöhte Scores zeigt, also bei den Konstellationen  $aab$  oder  $aba$  oder  $baa$ ; dafür errechnet sich die Zufallswahrscheinlichkeit ebenfalls zu  $3/8$ . Bei den insgesamt 15 Probanden wäre also unter Zufallsbedingungen (fehlendem Einfluss der Therapie, weder im hemmenden noch fördernden Sinne) ein Wert von 3 am wahrscheinlichsten in zwei Fällen erwarten, ebenso ein Wert von 0, während sich die Werte 1 und 2 mit etwa gleicher Häufigkeit ergeben sollten (etwa 5 bis 6mal). Tatsächlich ist der Wert 3 aber 5mal aufgetreten, was sich unter Zufallsbedingungen laut Binomialverteilung bestenfalls in 3% der Fälle beobachten lässt. Bestimmt man daraufhin noch die Wahrscheinlichkeit, dass von den verbleibenden 10 Personen mit einem Wert von weniger als 3 immerhin

vier den Minimalwert 0 zeigen, so berechnet sich die Zufallswahrscheinlichkeit dafür als etwa 4%. Somit lässt sich sagen: Die Therapie hat durchaus einen Einfluss gehabt, wenn auch einen für die einzelnen Teilnehmer deutlich unterschiedlichen. Mehr Personen als unter Zufallsbedingungen zu erwarten haben nach Abschluss der Therapie verstärkte Intensität sozialer Kontakte gezeigt; gleichfalls hat bei überzufällig vielen von ihnen die Intervention genau das Gegenteil bewirkt.

Anders als bei einer primären Datenaggregation konnte sekundäre Aggregation zunächst einzelfallanalytisch aufbereiteter Daten Effekte nachweisen (und zwar differenzielle). Zudem ließ sich jeder Proband mit einem zwar fehlerbehafteten, prinzipiell aber informativen Kennwert versehen, der seine Reaktion auf die Therapie anschaulicher beschreibt als die einfache Differenz zwischen seinem post- und seinem prä-Therapie-Score. Man könnte nun weiter gehen, nämlich Probanden mit wahrscheinlichen Therapierfolgen (Werte in Spalte 6 von 3) und solche mit wahrscheinlichen regelrechten Therapiemisserfolgen (Werte von 0) hinsichtlich diverser anderer Parameter vergleichen, um eventuelle Prädiktorvariablen für eine erfolgreiche Therapie zu finden.

Allerdings ist prinzipiell ebenso der Fall möglich, dass die aggregierten, einzelfallstatistisch abgesicherten Individualwerte sich weniger leicht gegen Zufälligkeit absichern lassen als die primär aggregierten Werte; ein Beispiel hierfür würde den gesetzten Rahmen jedoch sprengen.

Einzelfallanalytische Auswertung oder einzelfallanalytische Aufbereitung der Daten einer Probandenstichprobe schließt anschließende aggregatstatistische Auswertung keineswegs aus. Es ergibt sich sogar oft der Fall, dass hierdurch die Power des aggregatstatistischen Verfahrens erhöht wird, also die Chance eines signifikanten Resultats steigt. Einzelfallanalytische Auswertung oder Aufbereitung liefert zudem oft Hinweise, bei welchen Probanden sich Effekte nachweisen lassen und mit welchen Eigenschaften der Probanden dies assoziiert sein könnte.

## Anmerkungen zu Kapitel 1

1. Eine solche Untersuchungseinheit könnte beispielsweise eine Fußballmannschaft sein, wenn überprüft werden soll, ob sie in der zweiten Halbzeit generell leistungsmäßig nachlässt; ebenso wäre eine Untersuchungseinheit der kompliziert aus den Kursen von dreißig Aktienunternehmen berechnete Deutsche Aktienindex (DAX), wenn dessen Verlauf, etwa in Abhängigkeit vom Dollarkurs, charakterisiert wird. Im Rahmen psychologischer, pädagogischer oder medizinischer Fragestellungen sind es jedoch insbesondere Individuen, deren Daten studiert werden, und auf diesen Fall wollen wir uns im wesentlichen beschränken.

2. So einfach die Erhebung ist und so leicht sich die Interpretation zu gestalten scheint, stecken sie jedoch voller Tücken: So gibt es bei einmaligen Messungen keinen Anhalt für die individuellen Schwankungen des erhobenen Variablenwerts und die Größe des Messfehlers (z. B. die Beeinflussung des Blutdruckwerts durch den Erhebungsvorgang an sich); ebenso ist Ungenauigkeit oder Fehlerhaftigkeit des Messinstruments zu bedenken. Mittelung der Ergebnisse wiederholter Messungen bei diesem Probanden bietet ei-

ne Möglichkeit, seinem „wahren“ Wert in dieser Variable und dieser Situation wenigstens näher zu kommen.

3. Man beachte, dass in diesem Beispiel bereits ein einzelfallstatistischer Induktionsschluss vorliegt. Hätte die Aussage gelaute: Die durchschnittliche Zahl der täglichen sozialen Kontakte des Probanden liegt nach Therapie höher als vor Therapie, wären lediglich zwei Stichproben verglichen worden (Erhebungszeitpunkte vor und Erhebungszeitpunkte nach Therapie bei ein und demselben Individuum). Durch den Zusatz „signifikant“ ist gesagt, dass ein solcher Unterschied bei tatsächlicher Gleichheit der sozialen Kontakte in beiden Messsituationen nur in weniger als 5% der Fälle (nämlich von ähnlichen Erhebungen bei dem betrachteten Individuum) zu beobachten wäre; man ist deshalb umgekehrt gut beraten, nicht mehr von gleicher Anzahl der sozialen Kontakte vor und nach Therapie auszugehen.

4. Da zur Einzelfallanalyse generell eine gewisse Anzahl von Daten jedes Individuums erforderlich ist, lassen sich geeignete (beispielsweise alle in einer bestimmten Situation erhobene) Daten mitteln und sich der individuelle Fehlerwert vermindern, welcher zusammen mit dem „wahren Wert“ des Probanden in die Aggregatstatistik eingeht. Schon aus diesem Grund kann sich der Aufwand intensiverer Erhebung lohnen, auch wenn sich die Daten letztendlich doch nicht für einzelfallstatistische Behandlung eignen.

5. Es scheint nahe liegend, für jeden der Probanden aus den drei Messungen jeder Phase die Phasenmittelwerte und die Varianz innerhalb der Phase zu bilden und die Phasenunterschiede mittels des  $t$ -Tests (für unabhängige Stichproben!) auf Überzufälligkeit zu überprüfen; rein rechnerisch wäre dies möglich, die Ergebnisse aber nicht sicher zu interpretieren. Die Schwierigkeit ist hier, dass wir aus (vermutlich) seriell abhängigen Daten Kennwerte erhalten, deren Einsatz in die statistischen Formeln (beispielsweise für den  $t$ -Test) Unabhängigkeit der Ausgangsdaten verlangt (siehe 9.1).