

Teil III
**Multivariate
Methoden**

Einleitung

Die Beeinflussung einer abhängigen Variablen durch eine oder mehrere unabhängige Variablen kann auf vielfältige Weise mit den in Teil II besprochenen varianzanalytischen Methoden untersucht werden. In den Sozialwissenschaften befassen wir uns jedoch häufig mit abhängigen Variablen, die nur schwer und im allgemeinen nur unzureichend auf Grund eines einzelnen Indikators operationalisiert bzw. gemessen werden können. Dazu zählen beispielsweise komplexe Merkmale wie sozialer Status, berufliche Zufriedenheit, Therapieerfolg, Einstellungen, Begabungen, Interessen, Erziehungsstil, Krankheitssymptomatik usw. Die Erfassung dieser und ähnlich komplexer Merkmale durch nur einen Wert ist häufig unbefriedigend, und die statistische Analyse der an einer Stichprobe erhobenen Werte führt nicht selten deshalb zu widersprüchlichen Ergebnissen, weil in thematisch vergleichbaren Untersuchungen jeweils andere Teilaspekte des komplexen Merkmals herausgegriffen werden. Untersucht man komplexe Merkmale hingegen bezüglich vieler Teilindikatoren, führt dies zu stabileren und informationsreicheren Ergebnissen.

Mit multivariaten Methoden werden Hypothesen geprüft, die sich auf das Zusammenwirken vieler abhängiger und unabhängiger Variablen beziehen.

Die in Teil III unter dieser summarischen Beschreibung behandelten Verfahren seien im folgenden anhand einiger Fragestellungen vorgestellt.

Nach wie vor problematisch ist die Zuordnung von psychiatrischen Patienten zu einzelnen psychiatrischen Krankheitskategorien. Es soll deshalb überprüft werden, in welcher Weise sich Patienten, die als depressiv, schizophren, paranoid oder manisch klassifiziert wurden, hinsichtlich ihrer Krankheitssymptomatik unterscheiden.

Geschulte Psychiater werden gebeten, zufällig ausgewählte Patienten der 4 genannten Kategorien auf Ratingskalen danach einzustufen, wie stark die folgenden Merkmale ausgeprägt sind:

1. Angstgefühle
2. Denkstörungen
3. Feindseligkeit
4. Mißtrauen
5. hypochondrische Tendenzen
6. emotionale Labilität
7. motorische Verlangsamung
8. innere Erregung
9. Gedächtnisschwäche
10. Schuldgefühle
11. depressive Stimmungen
12. Desorientierung
13. ungewöhnliche Denkinhalte
14. halluzinatorisches Verhalten
15. affektive Stumpfheit
16. Manieriertheit.

Um vorurteilsfreie Ratings zu erhalten, werden die Psychiater über die erste Diagnose für die Patienten, nach der die Zuordnung zu den 4 o. g. Kategorien erfolgte, nicht informiert.

Formal geht es also in diesem Beispiel um die Beeinflussung mehrerer abhängiger Variablen durch eine unabhängige Variable. Für nur eine abhängige Variable (z. B. Angstgefühle) wäre die in Kap. 7 beschriebene einfaktorielle Varianzanalyse anzuwenden. Mit diesem Verfahren ließe sich überprüfen, ob und in welcher Weise sich depressive, schizophrene, paranoide und manische Patienten z. B. hinsichtlich der abhängigen Variablen „Angstgefühle“ unterscheiden. Das komplexe Merkmal „Krankheitssymptomatik“ wird jedoch nicht nur durch ein, sondern durch 16 Merkmale erfaßt, d. h. wir müßten 16 einfaktorielle Varianzanalysen durchführen, um die 4 Patientengruppen hinsichtlich der gesamten Krankheitssymptomatik differenzieren zu können.

Diese Vorgehensweise ist jedoch aus zweierlei Gründen nicht empfehlenswert. Erstens ist damit zu rechnen, daß aufgrund der einzelnen Varianzanalysen Entscheidungen getroffen werden, deren α - bzw. β -Fehler nur schwer kalkulierbar sind (vgl. hierzu S. 248 f.). Zweitens können sich die einzelnen Krankheitsmerkmale bei der Differenzierung der Krankheitsbilder gegenseitig ergänzen bzw. in kombinierter Form zu einer deutlicheren und besser interpretierbaren Unterscheidung der Gruppen führen als die 16 univariaten Analysen. Den 16 univariaten einfaktoriellen Varianzanalysen wäre deshalb ein multivariater Mittelwertsvergleich bzw. eine *multivariate Varianzanalyse* vorzuziehen. Soll die Bedeutung der erhobenen Merkmale für die Unterscheidung der 4 verschiedenen Gruppen genauer untersucht werden, wählen wir als Auswertungsverfahren eine *Diskriminanzanalyse* (vgl. Kap. 18).

Multivariate Mittelwertsvergleiche werden wir in Kap. 17 behandeln. Sie unterscheiden sich von univariaten Mittelwertsvergleichen (t-Test, univariate Varianzanalyse) darin, daß statt *einer* abhängigen Variablen *mehrere* abhängige Variablen simultan untersucht werden. Darüber hinaus besteht wie in der univariaten Varianzanalyse die Möglichkeit, die zu vergleichenden Vpn hinsichtlich mehrerer unabhängiger Variablen zu gruppieren. In diesem Falle sprechen wir von einer *mehrfaktoriellen, multivariaten Varianzanalyse*.

Kapitel 6 behandelte Methoden, mit denen der Zusammenhang zwischen zwei Merkmalen bestimmt bzw. eine Gleichung zur Vorhersage des einen Merkmals auf Grund des anderen erstellt werden kann (Korrelations- und Regressionsrechnung). Auch diese Verfahren werden im Teil III wieder aufgegriffen und zu einem allgemeinen, multivariaten Ansatz ausgebaut. Das folgende Beispiel verdeutlicht, was wir unter multivariaten Korrelationsmethoden verstehen wollen.

Ein Teilbereich der Psychologie, die physiologische Psychologie, untersucht u.a. die Frage, mit welchen physiologischen Variablen subjektiv empfundene Gefühle zusammenhängen. Mit den in Kap. 6 behandelten Verfahren könnten bivariate Korrelationen zwischen subjektiven Gefühlsvariablen (z. B. die vor einer Examens- oder Testsituation empfundene Angst) und physiologischen Variablen (wie z. B. die psychogalvanische Hautreaktion) ermittelt bzw. eine Regressionsgleichung zur Vorhersage von Angstgefühlen auf Grund der psychogalvanischen Hautreaktion (oder umgekehrt) aufgestellt werden. Der Literatur entneh-

men wir jedoch, daß nicht nur die psychogalvanische Hautreaktion, sondern eine Reihe weiterer physiologischer Variablen, wie z. B. Pulsfrequenz, Blutdruck, Pupillenöffnung, Flimmer-Verschmelzungsfrequenz, pH-Wert des Speichels, Blutzuckerspiegel usw. für Gefühlszustände bedeutsam seien. Wollen wir das physiologische Korrelat eines spezifischen Gefühlszustandes erfassen, reicht es sicherlich nicht aus, hierfür einzelne physiologische Indikatoren isoliert zu untersuchen; erfolgversprechender wäre eine Methode, die es gestattet, einen Gefühlszustand auf Grund des Zusammenwirkens möglichst vieler physiologischer Variablen zu beschreiben.

Sollen, wie im vorliegenden Beispiel, mehrere Prädiktorvariablen gleichzeitig mit einer Kriteriumsvariablen in Beziehung gesetzt werden, berechnen wir eine *multiple Korrelation*. Mit der *multiplen Regressionsrechnung* bestimmen wir eine Gleichung zur Vorhersage einer Kriteriumsvariablen bei gleichzeitiger Berücksichtigung mehrerer Prädiktorvariablen (Kap. 13).

Im oben genannten Beispiel bezeichneten wir die subjektiv erlebte Angst als Kriteriumsvariable und die physiologischen Variablen als Prädiktorvariablen, wobei diese Zuordnung jedoch nicht zwingend ist. Ähnlich wie in der bivariaten Regressionsrechnung ist es auch für die multiple Regressionsrechnung ohne Belang, welche Variablen als Prädiktor- und welche als Kriteriumsvariablen eingesetzt werden. Immer, wenn eine einzelne Variable mit mehreren anderen Variablen in Beziehung gesetzt werden soll, kommt die multiple Regressionsrechnung zur Anwendung.

Die Verallgemeinerung des multiplen Korrelations- und Regressionsansatzes läßt es zu, auch nominalskalierte Variablen als Prädiktorvariablen einzusetzen. Wir werden hierauf im Kap. 14 unter dem Stichwort „Das allgemeine lineare Modell“ eingehen. In diesem Kapitel wird gezeigt, daß die in Teil II behandelten varianzanalytischen Methoden Spezialfälle der multiplen Korrelations- und Regressionsrechnung sind.

Eine andere Problemsituation liegt vor, wenn die Bedeutung mehrerer Prädiktorvariablen für mehrere Kriteriumsvariablen von Interesse ist. In der oben erwähnten psychophysiologischen Untersuchung wäre beispielsweise zu prüfen, ob es sinnvoll ist, das subjektiv erlebte Angstgefühl nur durch eine einzelne Variable zu erfassen, oder ob ein aus mehreren Items bestehender Fragebogen, in dem neben der subjektiv erlebten Angst weitere Erlebnisinhalte wie z. B. sub-

jektive Leistungserwartung, Konzentrationsstörungen und Vitalität erfaßt werden, den Gefühlszustand umfassender charakterisieren. Soll die Bedeutung mehrerer Prädiktorvariablen für ein komplexes Kriterium, das ebenfalls durch mehrere Variablen operationalisiert ist, ermittelt werden, führen wir eine *kanonische Korrelationsanalyse* durch (Kap. 19).

Die kanonische Korrelation stellt eine Verallgemeinerung der multiplen Korrelation dar, die ihrerseits eine Verallgemeinerung der bivariaten Korrelation ist. Da Varianzanalysen im allgemeinen linearen Modell als Spezialfälle der multiplen Korrelations- und Regressionsanalyse aufgefaßt werden, sind sie natürlich auch Spezialfälle der kanonischen Korrelation. Unter 19.3 (Die kanonische Korrelation: Ein allgemeiner Lösungsansatz) werden wir zeigen, daß auch die multivariaten Mittelwertsvergleiche bzw. die Diskriminanzanalyse als kanonische Korrelationsanalyse darstellbar ist. Die Behauptung, die kanonische Korrelationsanalyse sei ein allgemeiner Lösungsansatz, ist schließlich dadurch zu rechtfertigen, daß wir auch die unter 5.3 behandelten Chi-Quadrat-Techniken als Sonderfälle der kanonischen Korrelation interpretieren können.

Die Zielvorstellung, ein komplexes Merkmal möglichst breit und differenziert erfassen zu wollen, ist häufig mit der Unsicherheit verbunden, welche Operationalisierungen des Merkmals in eine Untersuchung angesichts zeitlicher und finanzieller Rahmenbedingungen aufgenommen werden sollen. Dieses Problem führt zu der Frage, wie die Anzahl der zu erhebenden Variablen minimalisiert werden kann, ohne auf relevante Informationen zu verzichten. Ein unsystematisches Vorgehen bei der Variablenauswahl kann dazu führen, daß der Untersuchungsaufwand durch Variablen vergrößert wird, die nur redundante Informationen, d. h. Informationen, die bereits mit anderen Variablen erfaßt werden, liefern. Mit in diesem Sinne überflüssigen Informationen muß vor allem dann gerechnet werden, wenn die Variablen hoch miteinander korrelieren. So wissen wir beispielsweise, daß physiologische Indikatoren von Gefühlszuständen, von denen oben einige erwähnt wurden, wechselseitig korreliert sind. Es empfiehlt sich deshalb, eine Auswahl von physiologischen Variablen zu treffen, die wechselseitig möglichst wenig korreliert und damit nur wenig redundant sind.

Mit steigender Variablenanzahl kann die Variablenauswahl auf Grund der Interkorrelationen jedoch sehr bald nicht mehr zufriedenstellend vorgenommen wer-

den, weil die Anzahl der simultan zu berücksichtigenden Korrelationen zu groß wird. (Bei 10 Variablen müssen bereits 45 Korrelationen und bei 20 Variablen 190 Korrelationen gleichzeitig betrachtet werden.) Wir werden deshalb in Kap. 15 ein Verfahren kennenlernen, das die Zusammenhänge vieler Variablen analysiert und das damit entscheidend zur optimalen Variablenauswahl beitragen kann. Dieses Verfahren wird in seiner allgemeinen Form *Faktorenanalyse* genannt. (Im Rahmen der multivariaten Verfahren nimmt dieses Verfahren eine Sonderstellung ein, da es nicht zwischen abhängigen und unabhängigen Variablen unterscheidet.) Kapitel 16 behandelt als ein weiteres multivariates Verfahren die *Clusteranalyse*, die verwendet wird, um viele, multivariat beschriebene Untersuchungsobjekte in homogene Gruppen oder Cluster einzuteilen.

Multivariate Methoden gestatten die simultane Berücksichtigung sehr vieler Variablen, was zwangsläufig dazu führt, daß der mit diesen Verfahren verbundene rechnerische Aufwand weitaus größer ist als der Aufwand der bisher behandelten Verfahren. Der Einsatz multivariater Verfahren ist deshalb ohne die Benutzung elektronischer Datenverarbeitungsanlagen praktisch undenkbar. Zudem setzen multivariate Verfahren mehr mathematische Vorkenntnisse voraus als die bisher behandelten Verfahren, so daß der mathematisch weniger geschulte Leser eventuell einzelne Schritte nicht ohne weiteres nachvollziehen kann.

Es fragt sich allerdings, ob ein vollständiges Verständnis dieser zum Teil recht komplizierten Techniken überhaupt notwendig ist, denn zumindest die rechnerische Durchführung läßt sich mühelos mit einem statistischen Programmpaket erledigen. Wie in den vorangegangenen Kapiteln werden deshalb die SPSS-Ergebnisprotokolle der wichtigsten Verfahren im Anhang E dokumentiert und kurz interpretiert.

Nicht zu den Servicefunktionen der Anbieter von Statistik-Software zählt jedoch im allgemeinen die Vermittlung des Verständnisses eines Verfahrens, das notwendig ist, wenn ein Datenmaterial problemadäquat ausgewertet und die Ergebnisse richtig interpretiert werden sollen. Wir werden deshalb neben der ausführlichen, mit Beispielen versehenen Darstellung der Indikation und der Interpretation auch auf den Rechengang der Verfahren eingehen und damit dem Leser eine Möglichkeit anbieten, sich auch mit der Mathematik der Verfahren ein wenig vertraut zu machen.

Die Darstellung der Rechenregeln der multivariaten Verfahren wird durch den Einsatz der *Matrixalgebra* erleichtert. Wir empfehlen, sich vor dem Lesen derjenigen Teile, die sich mit dem mathematischen

Hintergrund der Verfahren bzw. ihrer rechnerischen Durchführung befassen, mit dem Grundprinzip der Matrixalgebra vertraut zu machen. Die dazu erforderlichen Regeln sind im Anhang C aufgeführt.