

Prof. Dr. Petra Stein
Universität Duisburg-Essen
Institut für Soziologie

Wintersemester 2011/2012

Multivariate statistische Analyseverfahren

MA Soziologie, MA Survey Methodology

Verfahren mit abhängigen und unabhängigen Variablen

Verfahren	Skalenniveau und Dimension q (abhängige Variablen)	Skalenniveau und Dimension p (unabhängige Variablen)
a) Lineare und nichtlineare multiple Regression	metrisch, $q = 1$	beliebig, $p \geq 1$
b) Multivariate Regression	metrisch, $q \geq 1$	beliebig, $p \geq 1$
c) Univariate Varianzanalyse	metrisch, $q = 1$	kategorial, $p \geq 1$
d) Multivariate Varianzanalyse	metrisch, $q \geq 1$	kategorial, $p \geq 1$
e) Kovarianzanalyse	metrisch, $q = 1$	kategorial/metrisch, $p \geq 2$
f) Kategoriale Regression	kategorial, $q \geq 1$	gemischt, $p \geq 1$
g) Diskriminanzanalyse	kategorial, $q \geq 1$	metrisch/kategorial/gemischt, $p \geq 1$
h) Logistische Regression	kategorial, $q = 1$	metrisch/kategorial, $p \geq 1$



Zielsetzungen der Verfahren

Analyse und Beschreibung des Einflusses:

Lineare und nichtlineare multiple Regression

(Effekts) von unabhängigen Variablen auf eine oder mehrere abhängige Variablen:

Multivariate Regression, Uni- und Multivariate Varianzanalyse, Kovarianzanalyse und Kategoriale Regression

Möglichst gute Unterscheidung der Kategorien einer nominalskalierten abhängigen Variablen aufgrund mehrerer (unabhängiger) Merkmale:

Diskriminanzanalyse



Beispiele zu Verfahren mit abhängigen und unabhängigen Variablen - 1

Für eine größere Gemeinde werden durch eine Stichprobenerhebung sogenannte Mitspiegel erstellt, um die ortsübliche Vergleichsmiete für „Wohnraum vergleichbarer Art, Größe, Ausstattung, Beschaffenheit und Lage“ zu schätzen. Sinngemäß bedeutet dies, dass die Nettomiete in Abhängigkeit von Merkmalen wie Wohnfläche, Alter der Wohnung bzw. des Gebäudes, Lage sowie Ausstattungsmerkmale (z.B. Zentralheizung) modelliert werden soll.



Beispiele zu Verfahren mit abhängigen und unabhängigen Variablen - 2

Der Marketingleiter eines Lebensmittelkonzerns steht vor der Frage, zu welchem empfohlenen Richtpreis eine neue Margarine-Marke auf den Markt gebracht werden soll. Neben der Preisstrategie ist auch die Wirkung des Faktors „Kommunikationsstrategie“ auf das Kaufverhalten von Interesse und zwar in den alternativen Formen „Postwurfsendungen“ und „Anzeigenwerbung“. Insgesamt sollen die Einflüsse der beiden Faktoren Preis und Kommunikationsstrategie einzeln und/oder in Kombination auf das Merkmal Absatz analysiert werden.



Beispiele zu Verfahren mit abhängigen und unabhängigen Variablen - 3

Einer Stichprobe von männlichen und weiblichen Ratten werden drei verschiedene Drogen verabreicht. Jeweils nach der ersten und zweiten Woche wird der Gewichtsverlust der Versuchstiere festgestellt. Es sollen die Einflüsse der Variablen „Geschlecht“ und „Drogenart“ auf die beiden Merkmale „Gewichtsverlust nach der ersten Woche“ und „Gewichtsverlust nach der zweiten Woche“ analysiert werden.



Beispiele zu Verfahren mit abhängigen und unabhängigen Variablen - 4

Die Vergabe von Privatkrediten wird von der „Bonität“ der Kunden abhängig gemacht, also von der Bereitschaft und Fähigkeit der Kunden, die anfallenden Zins- und Tilgungsraten ordnungsgemäß zu begleichen. Ein potentieller Kreditnehmer ist entweder der Klasse der problemlosen Kreditnehmer zuzuordnen oder in die Klasse der Problemfälle einzuordnen, bei denen eine genauere Prüfung vorzunehmen bzw. auf das Kreditgeschäft zu verzichten ist. Jeder Kreditkunde wird durch eine Reihe von Merkmalen charakterisiert. Auf der Basis dieser Merkmale wird versucht, eine statistisch gesicherte Entscheidung über Vergabe oder Ablehnung des Kredits zu treffen.



Verfahren mit symmetrischer Behandlung der Variablen

Verfahren	Skalierung der Merkmale	Zielsetzung
a) Diskriminanzanalyse	metrisch, kategorial, gemischt	Klassifikation von Objekten in vorgegebene Klassen
b) Clusteranalyse	beliebig	Zusammenfassung von Objekten zu homogenen Gruppen
c) Analyse mehrdimensionaler Kontingenztabellen	kategorial	Untersuchung der Zusammenhangsstruktur
d) Partielle, multiple und kanonische Korrelation	metrisch	Untersuchung der Zusammenhangsstruktur, Datenreduktion
e) Faktorenanalyse	(vorwiegend) metrisch	Möglichst gute Repräsentation von mehreren Variablen durch wenige hypothetische Variablen (Faktoren); Datenreduktion
f) Mehrdimensionale Skalierung	metrisch oder ordinal	Repräsentation ähnlicher Objekte in Räumen von möglichst niedriger Dimension



Beispiele zu Verfahren mit symmetrischer Behandlung der Variablen - 1

In einer Untersuchung zur Suizidgefährdung wurden 106 depressive Patienten, die während eines Jahres in eine Klinik eintraten, hinsichtlich des Auftretens von vier Symptomen (Stimmung gereizt, Stimmung gespannt, Denkablauf verlangsamt, Suizidgefährdung) beurteilt. Es sollen mögliche Zusammenhänge zwischen den vier Symptomen analysiert werden.



Beispiele zu Verfahren mit symmetrischer Behandlung der Variablen - 2

Um die Vorlieben ihrer Hörer zu analysieren, gibt eine Rundfunkgesellschaft eine Umfrage in Auftrag. Eine Anzahl von Personen wird befragt, welche Art von Programmen sie gerne hören, wobei eine Auswahl von bestimmten Programmtypen wie Nachrichten, Quiz-, Show-, und Pop-Sendungen, Sport, Serien usw. zur Verfügung steht. Mittels der Befragungsergebnisse sollen Gruppen „ähnlicher Hörer“, z.B. gewisse Typen von Durchschnittshörern, religiös-konservative Hörer usw., gebildet werden.



Beispiele zu Verfahren mit symmetrischer Behandlung der Variablen - 3

In einer Befragung von 114 Studenten des B.A. Soziologie sollen Gründe für die Wahl ihres Studienfaches untersucht werden. Die Studierenden können 13 Beweggründe (Variablen) mit einer Bewertung von eins bis sieben Punkten versehen und dadurch angeben, wie stark der jeweilige Aspekt in ihre Entscheidung einging. Ziel der Analyse ist eine Grobstrukturierung der 13 Variablen in wenige, trennbare Blöcke anhand einer Zurückführung der erhobenen Merkmale auf wesentliche Faktoren, wie etwa einem sozialen Faktor, sowie die Interpretation dieser Faktoren.



Beispiele zu Verfahren mit symmetrischer Behandlung der Variablen - 4

Zur Beurteilung und Entscheidung netzpolitischer Maßnahmen, insbesondere nach absatzrelevanten Kriterien, wurden über das Tankstellennetz der ARAL in der BRD im Rahmen einer Stichprobenziehung ($n = 700$) Informationen über Lage, Umgebung, Besitzverhältnisse, Ausstattung, Serviceangebot, Bedienung oder Selbstbedienung etc. erhoben. Als ein wichtiges Zielkriterium diente das Merkmal „Absatz“ mit den Kategorien „niedrig“, „mittel“ und „hoch“.



Basisliteratur I

- Fahrmeier, L., Hamerle, A. & Tutz, G. (1996): Multivariate statistische Verfahren. Berlin: de Gruyter. 2. Überarbeitete Auflage.
- Hartung, J., Elpelt, B. (2006): Multivariate Statistik. München: Oldenburg. 7. Unveränderte Auflage.
- Werner, J. (1997): Lineare Statistik: das allgemeine lineare Modell. Weinheim: Beltz.
- Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2005): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Berlin: Springer.



Basisliteratur II

- Bortz, J. (2005): Statistik für Human- und Sozialwissenschaftler. Heidelberg: Springer.
- Jann, B. (2002): Einführung in die Statistik. München: Oldenbourg.
 - Sarstedt, M., Schütz, T. (2006): SPSS Syntax. Eine anwendungsorientierte Einführung. München: Verlag Franz Vahlen.
 - Brosius, F. (2008): SPSS 16.0. Bonn: mitp-Verlag.

