

Zeitschrift: Schweizerische Zeitschrift für Soziologie = Revue suisse de sociologie
= Swiss journal of sociology

Herausgeber: Schweizerische Gesellschaft für Soziologie

Band: 11 (1985)

Heft: 3

Artikel: Zur Verletzung von Modellannahmen in der Regressionsanalyse

Autor: Sieber, M.

DOI: <https://doi.org/10.5169/seals-814964>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

Download PDF: 16.05.2025

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

ZUR VERLETZUNG VON MODELLANNAHMEN IN DER REGRESSIONSANALYSE *

M. Sieber

Psychiatrische Universitätsklinik Zürich,
Forschungsdirektion, Postfach 68, CH-8029 Zürich

1. Einleitung

Die Anwendung der Regressionsanalyse (RA) bei sozialwissenschaftlichen Daten ist nicht unbestritten, da häufig die Voraussetzungen für die Durchführung der RA nicht erfüllt sind. Gelangt die RA trotzdem zur Anwendung, so geschieht dies oft mit der Begründung, dass die Vorteile der RA die Nachteile überwiegen, und dass der Einfluss der nicht erfüllten Voraussetzungen die Resultate nur unwesentlich verändern würde. Dieses zweite Argument wird in der vorliegenden Arbeit näher untersucht. Im ersten Teil wird auf die theoretischen Aspekte der Voraussetzungsverletzungen eingegangen. Im nachfolgenden empirischen Teil wird bei einer früher durchgeführten RA geprüft, ob andere am gleichen Datensatz durchgeführte Analysen zu ähnlichen Ergebnissen führen.

2. Messtheoretische Voraussetzungen

Die nachfolgende Liste der Voraussetzungen bezieht sich auf das allgemeine regressionsanalytische Modell mit a priori festgelegten Werten der Prädiktoren (Gaensslen & Schuboe, 1973, 45-49). Zur Uebertragung auf das klassische korrelationsanalytische Modell (Modell mit stochastischen Prädiktoren) werden wir weiter unten zu sprechen kommen. Es gelten die folgenden Voraussetzungen:

* Diese Arbeit ist aus einem methodenkritischen Disput unter Fachkollegen entstanden, der in der Soziologie als "Undermeasurement-Kontroverse" bekannt ist. Ich danke F. Angst, R. Grünenfelder, R. Stähli und A. Tschopp für ihre kritischen Beiträge. Sie haben in der Kontroverse z.T. unterschiedliche Standpunkte eingenommen, was zum Ansporn des hier vorgelegten Diskussionsbeitrages wurde. Die finanzielle Unterstützung durch den Schweizerischen Nationalfonds (Kredit Nr. 3.887.0.81) sei dankbar erwähnt, ebenfalls die Mithilfe bei der Auswertung durch R. Grünenfelder.

- A1 - Die Beziehung zwischen den unabhängigen Variablen (UV), den Prädiktoren X , und der abhängigen Variable (AV) Y ist linear und in der einfachsten Form $Y = a + bX + e$, resp. expliziter für Person i :

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \dots \beta_m X_{mi} + \dots \beta_k X_{ki} + e_i$$
- A2 - Alle relevanten UV sind im Modell einbezogen (kein Spezifikationsfehler).
- A3 - Es darf keine vollständige lineare Beziehung zwischen den UV bestehen ($r < \pm 1.0$). (Multikollinearität).
- A4 - Die Zahl der Beobachtungen (P_{bn}) muss grösser sein als die Anzahl der UV.
- A5 - Die zu untersuchenden Rohwerte stellen Zufalls-Stichproben aus der interessierenden Grundgesamtheit dar. (Für Hypothesenprüfung).
- A6 - Die Variablen werden ohne Messfehler gemessen.
- A7 - Für den Erwartungswert des Vektors $\underline{e}$ gilt: $E_e(\underline{e}) = \underline{0}$.
- A8 - Die Fehlervariable hat innerhalb ausgewählter Werte von X homogene Varianz (Varianzhomogenität, Homoskedastizität).
- A9 - Sämtliche Fehlerwerte e sind voneinander unabhängig: $cov(e_i, e_h) = 0$; $i \neq h$, $i, h = 1, 2, \dots, N$. (Bei mehrfacher Wiederholung des Experimentes dürfte keine Kovarianz zwischen den Fehlerwerten etwa der beiden ersten beobachteten Personen (i, h) jeder der verschiedenen Stichproben auftreten).
- A10 - Die Fehlervariable e und die Prädiktoren X sind unkorreliert.
- A11 - Die Fehlerwerte e sind innerhalb ausgewählter Werte von X je normal verteilt. (Wichtig für Hypothesenprüfung).
- A12 - Obwohl ein bestimmtes Skalenniveau nicht zu den mathematischen Voraussetzungen zur Anwendung der multiplen RA (MRA) gehört, ist die MRA nur sinnvoll, wenn mindestens die abhängige Variable intervallskaliert ist.

Welches sind die Konsequenzen bei der Verletzung der einzelnen Voraussetzungen ? Wie kann die Verletzung erkannt und korrigiert werden ?

In A1 wird die Art der in der Regression beteiligten Grössen und ihr Zusammenhang eingeführt. Die darin enthaltene Linearitätsannahme besagt, dass die Y -Werte sich proportional zur Ver-

änderung der X-Werte verändern. Verhalten sich die empirischen Werte nicht linear zueinander, so können die Ergebnisse unangemessen sein. - Ob die Daten der Linearitätsannahme entsprechen, kann auf verschiedene Weise eruiert werden. Ein einfaches, übersichtliches Verfahren stellt die visuelle Residuenanalyse dar (Anscombe & Tukey, 1963 ; Draper & Smith, 1966). Bestehen nicht-lineare Zusammenhänge, so können die nicht-linearen Dependenzien evtl. in lineare transformiert werden (Urban, 1982, 165; Hager & Westermann, 1983, 114).

Bei Annahme A2 wird vorausgesetzt, dass alle relevanten und keine irrelevanten UV einbezogen sind. Was eine relevante Variable ist, wird vorwiegend aus theoretischen Überlegungen abzuleiten sein. Im Zweifelsfall wird eher eine UV zuviel als zuwenig in das Modell aufgenommen (Schuboe, Haagen & Oberhofer, 1983, 225). Das Auslassen relevanter UV führt zu verzerrten Schätzwerten. Der Spezifikationsfehler wird zwar nur bei wichtigen ausgelassenen Variablen und bei hoher Kovarianz bedeutend (Bohrnstedt & Carter, 1971, 128), stellt aber - wenn vorhanden - einen gravierenden Fehler dar. Er kann nur bedingt statistisch ermittelt werden und muss vor allem Gegenstand theoretischer Überlegungen sein.

Durch die Annahmen A3 und A4 wird lediglich der unbequeme Fall der Nicht-Identifizierbarkeit von b ausgeschlossen. Die Forderungen können durch Umorganisation oder Verkleinerung der UV erfüllt werden.

In A5 wird auf die zufällige Auswahl der Personen hingewiesen (Gaensslen & Schuboe, 1973, 45). Parametrische Testverfahren können aber auch dann durchgeführt werden, wenn die beobachteten Eiheiten keine Zufallsstichprobe aus einer real existierenden Population darstellen (Hager & Westermann, 1983, 109).

Die Voraussetzung A6, dass die UV ohne Fehler gemessen werden, ist wohl im regressions-, nicht aber im korrelationsanalytischen Modell erfüllt, weshalb man sich dort mit der Annahme kleiner relativer Messfehler zufrieden geben muss (siehe unten).

In A7 wird ausgesagt, dass für jede Stufe der UV der Erwartungswert der Fehlervariable e Null ist. Trifft dies nicht zu, so ist die Schätzung des a -Koeffizienten verzerrt. Da die Konstante in der Sozialforschung oft von geringem theoretischem Interesse ist, so ist diese Fehlerquelle im allgemeinen unbedeutend (Lewis-Beck, 1980, 28; Urban, 1982, 160).

Gelangt das OLS-Parameter-Schätzverfahren zur Anwendung, so wird zudem Varianzhomogenität (A8) und Unkorreliertheit der

Fehler (A9, A10) vorausgesetzt (Schuboe et al., 1983, 216). Ist die Varianzhomogenität nicht gegeben, kann die OLS-Schätzmethode auch keine effizienten Resultate liefern (verzerrte Varianz von b), was Signifikanztests in Frage stellt. Zur Ermittlung von Streuungsungleichheit kann u.a. wiederum die visuelle Residuenanalyse eingesetzt werden. Mittels modifizierter Schätzmethoden (Weighted OLS-Methode) können die Effekte abgeschwächt werden (Urban, 1982, 196). Evtl. muss auch auf Signifikanztests verzichtet werden. Die Schätzer a und b bleiben jedoch auch bei Varianzheterogenität unverzerrt. Die Auswirkungen der Varianzheterogenität auf das Verhalten parametrischer Tests ist sehr ausführlich untersucht worden. Gemäss der Zusammenfassung bei Hager & Westermann (1983, 111) lässt sich entnehmen, dass in den meisten Fällen Varianzheterogenität hinsichtlich der Fehler erster und zweiter Art tolerabel ist, wenn die Stichprobengrössen für alle Bedingungen gleich sind. Liegen jedoch ungleiche Stichprobenumfänge vor, wird von der Verwendung der F - und t -Statistiken abgeraten. Im Falle ungleicher, jedoch grosser Stichproben können die bei Horton (1978, 42) aufgeführten Tabellen zur Schätzung des tatsächlichen Fehlers erster Art herangezogen werden.

Annahme A9 (Autokorrelation) ist ausser bei Zeitreinanalysen kaum relevant, da i.a. davon ausgegangen werden kann, dass sich die Pbn nicht gegenseitig beeinflussen. Ist die Annahme nicht erfüllt, so kann evtl. eine Neuspezifikation oder die doppelstufige WLS-Methode weiterhelfen. Die Annahme kann bedingt durch die visuelle Residuenanalyse oder den Durbin-Watson-Test geprüft werden (Urban, 1982, 217). Gegebenenfalls muss auf Signifikanztests verzichtet werden.

Annahme A10 betrifft die Unabhängigkeit der Fehler von den UV. Die Verletzung von A10 kann wiederum mittels visueller Residuenanalyse oder Korrelationsberechnungen zwischen Residuen und UV ermittelt werden. Je nachdem, welche Ursachen für die Verletzung von A10 in Frage kommen, sind auch die Folgen unterschiedlich: Bei fehlender Linearität sind a und b verzerrt, und R^2 wird zu klein geschätzt. Liegt ein Spezifikationsfehler vor, ist die Signifikanzprüfung des verzerrten b problematisch. Handelt es sich um ein Multikollinearitätsproblem, könnten signifikante Ergebnisse kaum entdeckt werden (grosser Standardfehler). Die Multikollinearität wird allerdings erst bei hoher Prädiktorenkorrelation relevant. Korrekturmöglichkeit für diese drei Ursachen sind bei Urban (1982) aufgeführt.

Bei A11 wird eine mehrdimensionale Normalverteilung der Fehlerwerte gefordert. Diese Annahme ist im Hinblick auf Signifikanztests wichtig und als solche schwer zu überprüfen. Bei grossen Stichproben und korrekter Spezifikation werden jedoch

die Residuen mit guter Näherung als multivariat normalverteilt betrachtet (zentraler Grenzwertsatz; Schuboe et al., 1983, 229; Urban, 1982, 158).¹ Betroffen wird wiederum die Robustheit der Konfidenzintervalle, was Signifikanztests einschränkt. Die Schätzer a und b bleiben unverzerrt. Zur Frage der Robustheit der t -, χ^2 - und F -Statistiken bei Normalitätsverletzung sind zahlreiche Untersuchungen durchgeführt worden. Sie zeigen, dass diese Tests über einen weiten Bereich der Abweichungen von der Normalität robust sind (Hager & Westermann, 1983, 110). Im einzelnen ist mit Ueberschreitungen der nominellen Fehlerwahrscheinlichkeit bei breitgipfligen oder L-förmigen Verteilungen zu rechnen. Mit Unterschreitungen muss man bei spitzgipfligen Verteilungen rechnen; Abweichungen von der Symmetrie wirken sich vor allem bei einseitigen Signifikanztests aus. Aus den Tabellen bei Horton (1978, 50) können die aktuellen Fehlerwahrscheinlichkeiten für verschiedene Verteilungen abgelesen werden.

Bleibt in der Aufstellung noch Annahme A12, wonach die AV Intervallskalenniveau aufweisen soll (für UV nicht erforderlich). Strittiger Punkt ist dabei die Frage, ob auch Ordinalskalen verwendet werden dürfen resp. welche Verzerrungen dabei impliziert werden. In der Psychometrie ist diese Kontroverse unter der Bezeichnung "Strong Statistic" versus "Weak Measurement" Position und in der Soziologie als "Undermeasurement-Kontroverse" bekannt. Das Problem wurde aus verschiedenen Gesichtspunkten betrachtet. Mittels Simulationsstudien wurde der Effekt unterschiedlicher Bewertungssysteme untersucht (Labovitz, 1967, 1970; Baker et al., 1966; Bollen & Barb, 1981) und kritisiert (Mayer, 1970, 1971). Allerbeck (1978) und Kim (1976) analysierten die Unterschiede bei der Verwendung von r resp. Kendall-tau. Theoretische Aspekte diskutieren Allerbeck (1978) und Baker et al. (1966) sowie Hager & Westermann (1983; "Messung per Fiat"). Lantermann (1976) weist auf den Nützlichkeitsaspekt hin. Der Grundtenor dieser hier zitierten Arbeiten geht dahin, dass in zahlreichen Fällen nur geringe Änderungen bei den Ergebnissen entstehen, wenn man ordinale Messungen als intervallskaliert betrachtet und dass die Vorteile dieses Vorgehens die Nachteile überwiegen. Voraussetzung ist aber, dass zwischen dem Ordinalmass und der wahren Skala eine monotone Beziehung besteht.

Die Bedeutung der Verletzungen der aufgeführten Annahmen ist unterschiedlich. Gravierender wird im allgemeinen das Mess-

¹ Bohrnstedt & Carter (1971, 131) sowie Wottawa (1982, 258) weisen darauf hin, dass nicht, wie irrtümlich manchmal postuliert, eine multivariate Normalverteilung der Variablen vorliegen muss, sondern dass für die bedingten Verteilungen der Fehlerwerte Normalität gefordert wird. Ueber die Verteilung der UV in der Population müssen keine Einschränkungen gemacht werden.

fehler- und Spezifikationsproblem erachtet, ferner auch die Korrelation zwischen UV und Messfehler (Johnston, 1972, 279; Lewis-Beck, 1980, 30). Während verschiedene Autoren die Robustheit gegenüber Verletzungen der Varianzhomogenität und Normalverteilung der Residuen betonen, nimmt Bradley (1978, 1980) eine andere Position ein. Da komplexe Versuchspläne mit verschiedenen Kombinationen von Voraussetzungsverletzungen kaum vorliegen, kann die Robustheit in solchen Fällen noch nicht beurteilt werden.

Alle aufgeführten 12 Annahmen beziehen sich wie erwähnt auf das klassische regressionsanalytische Modell mit fixierten Werten auf den UV. Wie steht es nun mit der Uebertragung auf das korrelationsanalytische Modell, wo die UV selbst Zufallsvariablen sind? Unter den aufgeführten, leicht modifizierten Annahmen (Schuboe et al., 1983, 233) behalten alle Ergebnisse bis auf eine Annahme ihre Gültigkeit. Die Schätzer bleiben erwartungstreu und Bereichsschätzungen sind möglich, obwohl für die Verteilungsform der UV keine Aussagen gemacht werden (Schuboe et al., 1983, 234; Johnston, 1972, 30). Die Ausnahme betrifft die b-Koeffizienten, welche ihre Blue-Eigenschaft² verlieren, jedoch asymptotisch effizient sind. Besonders problematisch wird nun aber die Annahme A6, wonach die UV fehlerfrei gemessen sind. Bei Schuboe et al., (1983, 239), Hartung (1982, 601) oder Schönfeld (1971, Kap. 11) wird angegeben, unter welchen zusätzlichen Annahmen (die Fehler betreffend) die Parameter geschätzt werden können. Es wird eine gewisse a priori-Kenntnis über die Varianzen der Fehler in den Variablen benötigt, um konsistente Parameter schätzen zu können. Steht diese Kenntnis nicht zur Verfügung, so muss man von der Annahme relativ kleiner Messfehler ausgehen, inferenzstatistische Schlüsse bleiben aber dann nur noch näherungsweise gültig (Schuboe et al., 1983, 250).

An dieser Stelle wird wohl deutlich, dass eine Ueberprüfung all dieser Annahmen im korrelationsanalytischen Modell noch schwieriger wird als im regressionsanalytischen. Mag es in der Praxis noch möglich sein, die Annahmen einzeln zu prüfen, so bleibt es doch unklar, wie sich die Effekte kumulativ auswirken, wenn mehrere Annahmen nicht erfüllt sind. Empirisch abgesicherte Faustregeln gibt es nicht. Was bleibt ist ein vorsichtiges Abwägen der Vor- und Nachteile unter Berücksichtigung der möglichen Voraussetzungsverletzungen beim zu analysierenden

² Ein Schätzwert hat Blue-Eigenschaften, wenn er ein "best, linear unbiased estimator" ist. Mit der OLS-Schätzung kann man die besten linearen und unverzerrten Schätzwerte für die Regressionskoeffizienten erreichen. Sie sind effizient und konsistent. Voraussetzung dafür ist allerdings die Gültigkeit der vier Annahmen A7 - A10.

Datensatz. Letzteres wird entscheidend sein für die Wahl der Interpretationsebene (mit oder ohne Signifikanzberechnung). Die Prüfung der einzelnen Voraussetzungen und die Durchführung von Plausibilitätsstudien geben wichtige Entscheidungsgrundlagen für die Beurteilung der Angemessenheit der MRA bei einem bestimmten Datensatz.

3. Ueberprüfung am konkreten Datensatz

Im folgenden soll an einem konkreten Datensatz geprüft werden, welche Voraussetzungen für die Durchführung der MRA nicht erfüllt wurden und ob die Ergebnisse mit anderen Berechnungen reproduziert werden können. Beim Datensatz handelt es sich um eine Studie zur Bedingungsanalyse des Konsums illegaler Drogen. Von 841 Männern lagen verschiedene Angaben zum sozialen Hintergrund, zur Persönlichkeit sowie zu Einstellungs- und Beziehungsaspekten vor, wobei einzelne, theorieorientierte Teilaspekte sowie ein zusammenfassendes Gesamtmodell mittels MRA analysiert wurden (Sieber & Angst, 1981). Im Gesamtmodell, welches hier überprüft wird, wurde der Konsum illegaler Drogen mit einer vierstufigen Rangskala erfasst und ging als AV in eine MRA mit 37 Prädiktoren ein.³ Bei der Interpretation der UV wurde die Höhe der Betagewichte wie auch deren Signifikanzniveau herangezogen.

Als erstes wurde die Voraussetzung homogener Fehlervarianzen (A8) geprüft, welche für den vorliegenden Datensatz nicht erfüllt war. Dies geht aus der visuellen Residuenanalyse (Streudiagramm der Residuen mit $\hat{Y}$) und aus dem Varianzvergleich der Residuen verschiedener Teilgruppen ($p < 0.01$) hervor. Benützt man die von Horton (1978) aufgeführten Tabellen zur groben Schätzung des Verzerrungseffektes, so resultiert aus dem Varianzvergleich eine aktuelle Irrtumswahrscheinlichkeit für den Alphafehler von 8 % anstelle des nominellen Fehlers von 5 %.

Die Annahme der Unabhängigkeit der Residuen von den Prädiktoren (A10) wurde zuerst anhand der Korrelation zwischen den Residuen und dem modellspezifischen Schätzer $\hat{Y}$ ermittelt (Draper & Smith, 1966, Kap. 3; Urban, 1982, 180). Sie muss

³ Das Merkmal ist extrem schief verteilt: $M = 1.803$, $S = 0.897$, Schiefe = 0.895, Exzess = -0.087. Die vier Ausprägungen lauten: Abstinente $N = 387$, Neugierkonsumenten $N = 283$, mittlere Konsumenten = 121, starke Konsumenten $N = 50$. Eine Ueberprüfung mit einer Guttman-Skalierung ergab eine sehr gute Skalierbarkeit. Dies deutet darauf hin, dass der illegale Drogenkonsum als rein quantitative, ordinale Grösse aufgefasst werden kann.

ebenfalls zurückgewiesen werden, da ein signifikanter Zusammenhang vorhanden war ($\tau\text{-}b = -.11$, $p = 0.001$). Bei den Korrelationen zwischen den Residuen und den einzelnen UV wurde bei keiner der 37 Prädiktoren ein signifikanter Zusammenhang gefunden. Gestützt auf die Simulationsstudie von Bohrnstedt & Carter (1971) wäre anzunehmen, dass die erwähnte schwache Korrelation von $-.11$ eher einen geringfügigen Fehler des Betakoeffizienten verursacht.

Die visuelle Residuenanalyse (Streudiagramm zwischen Residuen und Y-Werten sowie Residuen und X_k -Werten, (standardisierte Werte) erlaubt auch, auffällige Verletzungen der Linearitätsannahme (A1) zu entdecken. Für unseren Datensatz konnten wir keine derartigen Abweichungen feststellen, die Streudiagramme entsprachen aber auch nicht dem Idealfall (zufällige Streuung der Residuen um die Nullachse), sondern eher einem Bild, bei welchem ein Spezifikationsfehler vorliegen könnte.

Zur Entdeckung vorhandener Multikollinearität (A3) kann die lineare Abhängigkeit durch die Regression einer jeden X-Variablen auf jeweils alle anderen X-Variablen getestet werden (Urban, 1982, 187). Mehrere solche MRA an unserem Datensatz ergaben keine Hinweise auf Multikollinearität.

Zur Prüfung der Autokorrelation (A9) verwendeten wir den Durbin-Watson-Test (1971), welcher keine Verletzung von Annahme A9 anzeigte.

Als letztes wurde die visuelle Residuenanalyse im Hinblick auf die Normalverteilungsannahme (A11) herangezogen. Sie ergab, dass bei einigen der 37 Streudiagramme zwischen den Residuen und den einzelnen UV die Normalverteilungsannahme nicht erfüllt war.

Berücksichtigt man bei dieser Aufstellung auch noch die nicht erfüllte Annahme fehlerfreier Messungen (A6), so ist ersichtlich, dass mehrere Voraussetzungen im vorliegenden Datensatz nicht erfüllt sind und die Anwendung von Signifikanztests von daher nicht gerechtfertigt ist. Unklar bleibt die Frage, in welchem Ausmass auch die Betagewichte verzerrt sind. Im nun folgenden letzten Teil werden deshalb die Ergebnisse verschiedener Zusatzanalysen vorgestellt, bei welchen die Reproduzierbarkeit der Ergebnisse untersucht wurden.

In Abb. 1.1. (ausgezogene Linie) sind die Resultate der Original-MRA dargestellt. Die dazugehörigen Betakoeffizienten sind hinter der Variablenbezeichnung aufgeführt. Die UV sind in absteigender Reihenfolge der Betagewichte (Absolutwerte) geordnet.

Die ersten 11 Prädiktoren weisen signifikante Betagewichte auf ($p = .05$), die für Vergleichszwecke links der Kurve aufgeführt sind.⁴

Ergebnisse verschiedener Zusatzanalysen zur Original-Regressionsanalyse in Abb. 1.1: Beta-Gewichte für 6-stufige abhängige Variable (Abb. 1.2); Stichprobenhalbierung (Abb. 2.1/2); Diskriminanzanalyse (Abb. 3).

3.1. Rangkorrelationen

Als erste Zusatzanalyse wurde eine MRA mit einer Rangkorrelation-Eingabematrix (Kendall tau-b) berechnet. Erwartungsgemäss war die erklärte Varianz mit $R^2 = 0.36$ niedriger als in der Original-MRA mittels Produkt-Moment-Korrelationen ($R^2 = 0.44$).⁵ Der Profilverlauf der Beta-Koeffizienten ist demjenigen in Abb. 1.1 sehr ähnlich. Betrachtet man die UV mit signifikanten Beta-Koeffizienten ($p < 0.05$) als "bedeutungsvolle" Prädiktoren, so ergibt sich das identische Ergebnis wie in Abb. 1.1. Die Verwendung der den Daten angemesseneren Rangkorrelationen führen zu keinen anderen Interpretationen.

3.2. Abstufung der abhängigen Variablen

Bollen & Barb (1981, 232) zeigten, dass die Verwendung parametrischer Verfahren bei lediglich ordinalskalierten Variablen zu geringeren Verzerrungen führt, wenn die Ordinalskala in mindestens 5 Kategorien aufgeteilt wird. Dieser Aspekt wie auch die erwähnte Verletzung der Normalitätsannahme (A11) führte uns dazu, aus der 4-stufigen eine 6-stufige Kriteriumsvariable zu konstruieren. Wir teilten dazu die Abstinente illegaler Drogen zusätzlich in folgende drei Gruppen auf: 1. Abstinente mit zusätzlicher Tabak- und Alkoholabstinenz, 2. mit Zigarettenabstinenz, 3. restliche Abstinente illegaler Drogen. Damit ergab sich eine feiner abgestufte, bessere Verteilung der AV mit Gruppengrössen von 19, 89, 279, 283, 121 und 50 ($N = 841$). Das Ergebnis dieser MRA ($R^2 = 0.42$) ist in Abb. 1.2 (unterbrochene Linie) dargestellt. Der Profilverlauf entspricht etwa demjenigen der Original-MRA in Abb. 1.1 und alle signifikanten UV in 1.1

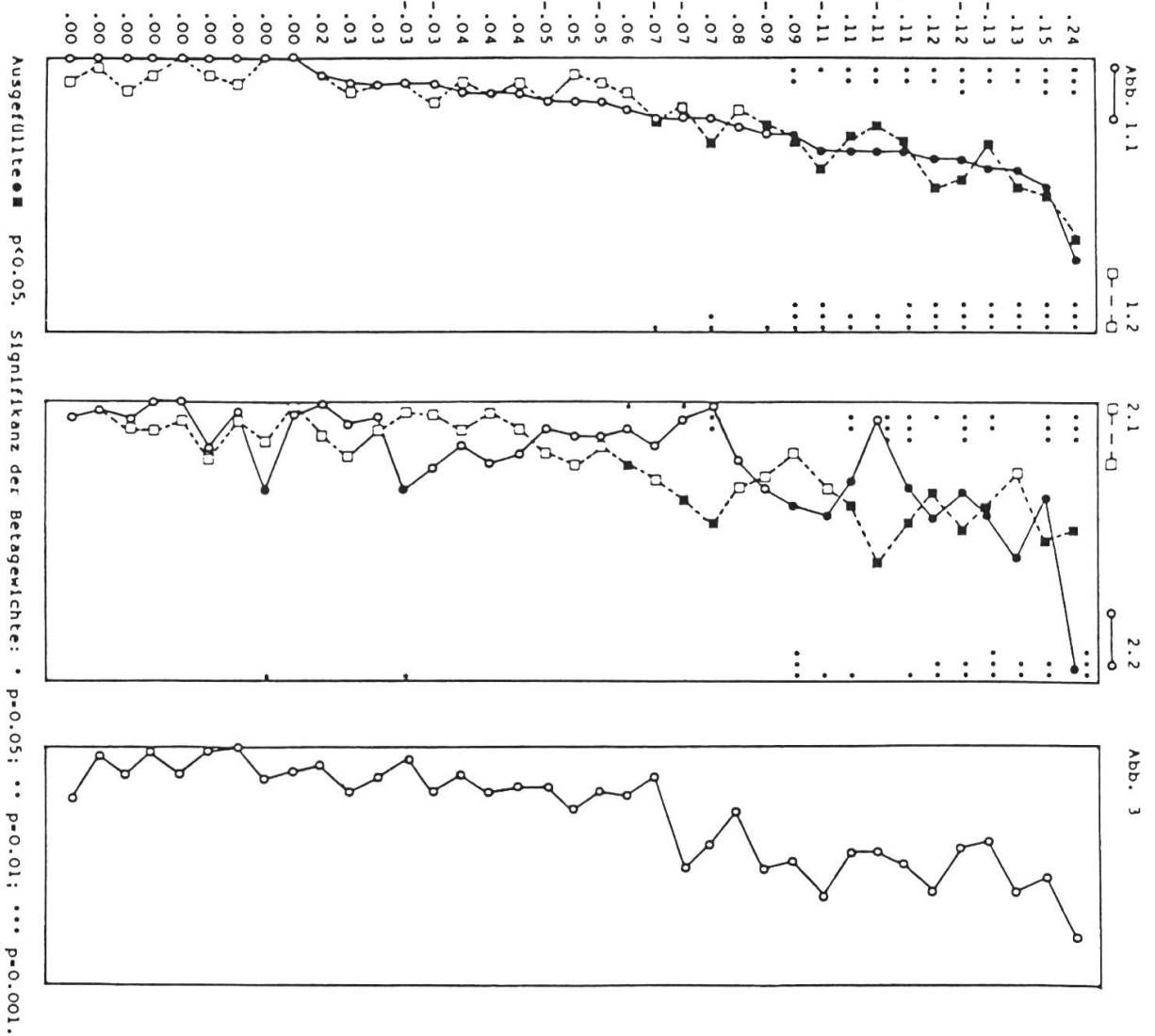
⁴ MRA mit paarweisen Ausschluss fehlender Daten. Ergänzende Analysen wurden mit allen 841 Pbn und mit der SPSS-Option "listwise deletion" ($N = 456$) durchgeführt. Sie führten zu einem fast identischen Kurvenverlauf wie in Abb. 1.1.

⁵ Ein Vergleich der Korrelationsmatrizen ergab, dass die Produkt-Moment-Korrelationen die "wahren" Korrelationen leicht überschätzen. Die Mehrzahl der Abweichungen lag zwischen 0.02 und 0.05, die grösste Abweichung betrug 0.13.

Abbildung 1 - 3:

1	Auslandaufenthalte	.24	...
2	Kirchenbesuche	.15	...
3	Depressivität	.13	...
4	Einstellung zu Randgrp.	-.13	...
5	Cliquenzugehörigkeit	-.12	...
6	Nervosität	.12	...
7	Einstellung zu Integr. Grp.	.11	...
8	Schwierigkeiten m. Polizei	-.11	...
9	Psychiatr. Behandlung	.11	...
10	Gehemtheit	-.11	...
11	Konfession	.09	...
12	Erregbarkeit	-.09	...
13	Schulbildung	.08	...
14	Geselligkeit	.07	...
15	Dominanzstreben	-.07	...
16	Schulbildung Mutter	-.07	...
17	Emot. Beziehung zu peers	.06	...
18	Sport in Freizeit	-.05	...
19	Aggressivität	.05	...
20	Berufsbildung	-.05	...
21	Berufstatus Vater	.04	...
22	Broken home	.04	...
23	Emot. Beziehung zu Eltern	.04	...
24	Berufliche Zufriedenheit	-.03	...
25	Alkoholrüsche Vater	-.03	...
26	Kontakte zur Mutter	.03	...
27	Regionale Herkunft	.03	...
28	Schulbildung Vater	.02	...
29	Ausbildung abgebrochen	.00	...
30	Alkoholrüsche Mutter	.00	...
31	Bewertung der Erziehung	.00	...
32	Berufswechsel	.00	...
33	Unregelm. Beschäftigung	.00	...
34	Beruflicher Aufstieg	.00	...
35	Gesundheitsverhalten	.00	...
36	Gelassenheit	.00	...
37	Offenheit	.00	...

Erläuterungen siehe Text.



sind auch in 1.2 signifikant. Zusätzlich sind 3 weitere signifikante UV dazugekommen. Diese ergänzende MRA bestätigt somit die Ergebnisse der Original-MRA und weist auf 3 weitere Prädiktoren hin, die von Bedeutung sein könnten.

3.3. Dichotomisierte Prädiktoren

In einer weiteren Analyse wurde der Effekt der unterschiedlichen Zahl der Abstufung innerhalb der 37 UV geprüft. Dazu wurden alle UV dichotomisiert. Das Ergebnis bestätigt im wesentlichen die Original-MRA in Abb. 1.1. Mit Ausnahme der "psychiatrischen Behandlungen" waren alle 11 in Abb. 1.1 signifikanten UV ebenfalls signifikant ($R^2 = 0.38$). Bei 6 weiteren Variablen sind neu signifikante Ergebnisse aufgetreten. Trotz des Informationsverlustes bei der Dichotomisierung konnten die bedeutungsvollen Prädiktoren der Original-MRA bestätigt werden.

3.4. Stichprobenhalbierung

Zur Prüfung der Stichprobenabhängigkeit der Betakoeffizienten halbierten wir den Datensatz in zwei Teilstichproben A und B und berechneten separat je eine MRA. Die Ergebnisse zeigen, dass die Anzahl der signifikanten Prädiktoren erwartungsgemäss abnimmt. In Teilstichprobe A (Abb. 2.1) waren 8 und in Stichprobe B (Abb. 2.2) 10 der 11 signifikanten UV der Original-MRA ebenfalls signifikant. ($R^2 = 0.46$, $N = 426$ resp. $R^2 = 0.47$, $N = 415$). Betrachtet man lediglich diejenigen UV als relevant, welche bei *beiden* Stichproben signifikante Betakoeffizienten aufweisen, so sind 7 der 11 Prädiktoren bedeutungsvoll, d.h. rund 2/3 der signifikanten Prädiktoren haben sich hier als stichprobenunabhängig erwiesen.

3.5. Kreuzvalidierung

Eine andere Methode zur Prüfung der Stichprobenabhängigkeit stellt die Kreuzvalidierung dar. Wir verwendeten dazu die unstandardisierten Regressionskoeffizienten der MRA in Teilstichprobe A und berechneten mittels Regressionsgleichung anhand der Daten von Stichprobe B die geschätzten $\hat{Y}$ -Werte für Stichprobe B. Der Mittelwert von $\hat{Y}$ betrug 1.77 ($S = 0.64$, $N = 415$) und lag etwas tiefer als der tatsächliche Mittelwert der AV in Stichprobe B, welcher 1.82 betrug ($S = 0.63$). Der für Stichprobe B aufgrund der Daten von Stichprobe A vorhergesagte Drogenkonsum wird deshalb im Mittel geringfügig unterschätzt. Die Korrelation des

Schätzwertes $\hat{Y}$ und dem tatsächlichen Y-Wert von Stichprobe B betrug $r = 0.59$, was einer erklärten Varianz von $R^2 = 0.35$ entspricht. Vergleichen wir diesen Wert mit demjenigen, der aus der MRA von Stichprobe B resultiert ($R^2 = 0.47$), so stellen wir fest, dass die Vorhersage aufgrund der Stichprobe A erwartungsgemäss niedriger, aber nicht wesentlich schlechter ausgefallen ist als die Vorhersage mittels direkt berechneter MRA. Die Beta-gewichte der ersten Teilstichprobe sind demnach insgesamt betrachtet brauchbare Schätzwerte zur Vorhersage des Zielkriteriums in der zweiten Teilstichprobe.

3.6. Diskriminanzanalyse

Da die Diskriminanzanalyse (DA) lediglich nominales Skalenniveau für die AV verlangt, gilt sie als ein voraussetzungsärmeres Verfahren als die MRA. Das Ergebnis der DA mit der 4-stufigen Kriteriumsvariable ist in Abb. 3 dargestellt. Im Profil sind die standardisierten Koeffizienten der ersten Diskriminanzfunktion aufgezeichnet.⁶ Der Profilverlauf zeigt, dass eine Aehnlichkeit zur Original-MRA, aber auch zur MRA mit einer 6-stufigen Kriteriumsvariable (Abb. 1.2) besteht. Nach der 15. UV "Dominanzstreben" sinkt das Profil deutlich ab. Die Bedeutsamkeit der 11 signifikanten UV der Original-MRA werden durch die DA bestätigt. Im weiteren scheinen gemäss DA aber auch die Prädiktoren Nr. 12, 14 und 15 wichtig zu sein, was mit der "besseren" MRA in Abb. 1.2 korrespondiert.

3.7. Multiple Klassifikationsanalyse (MCA)

In der letzten Analyse wurde eine MCA durchgeführt, welche infolge Speicherplatzkapazität auf 32 UV beschränkt werden musste. Wir überprüften hier, welche der 11 in Abb. 1.1 signifikanten UV auch in der MCA zu den wichtigeren Prädiktoren gehörten.⁷ Acht der 11 signifikanten UV erfüllten dieses Kriterium, d.h. die Mehrzahl der wichtigen Prädiktoren der Original-

⁶ Die DA ergab zwei signifikante Diskriminanzfunktionen (χ^2 -Test), wobei die erste mit einer kanonischen Korrelation von 0.66 84 % der aufgeklärten Varianz erklärte und damit wesentlich bedeutsamer war als die zweite Diskriminanzfunktion. Die Prognose der Gruppenzugehörigkeit gelang unterschiedlich gut: Von den 387 Abstinenden wurden 69.8 % richtig klassiert, gefolgt von den schweren Konsumenten mit 60 %. Bei Gruppe 2 wurden 49.1 % und bei Gruppe 3 37.2 % korrekt klassiert. Insgesamt wurden 57.5 % der Pbn richtig vorhergesagt (allein durch Zufall 35 %).

⁷ Als wichtiger Prädiktor wurden bei der MCA Prädiktoren mit Betagewicht < 0.09 (= Median) operational definiert.

MRA konnten, durch die MCA bestätigt werden.⁸ ($R^2 = 0.52$, adjustiertes $R^2 = 0.40$). Die drei nicht bestätigten Prädiktoren waren Cliquenzugehörigkeit, Gehemmtheit und Konfessionszugehörigkeit. Zusammenfassend resultiert aus diesen Zusatzberechnungen, dass die 11 "bedeutungsvollen" Prädiktoren ($p < 0.05$) der Original-MRA mehrheitlich bestätigt werden: Bei zwei Zusatzanalysen sind alle 11, bei je einer Zusatzberechnung sind 10, resp. 8, resp. 7 der 11 signifikanten Prädiktoren bestätigt worden.

4. Diskussion

Bedenkt man die zahlreichen Voraussetzungen, die für die Durchführung der MRA erfüllt sein müssen, sowie die Schwierigkeiten bei der Ueberprüfung dieser Voraussetzungen, so könnte man geneigt sein, überhaupt keine multiplen Regressionsanalysen durchzuführen. Der Forscher würde sich dabei auch die Kritik aus dem Lager der theoretisch orientierten Statistiker ersparen, ein Verfahren angewendet zu haben, das nicht zulässig ist. Da jedoch bei der Anwendung alternativer Verfahren häufig auch Voraussetzungen verletzt werden und/oder andere Nachteile auftreten (z.B. Informationsverlust), stellt sich die Frage, in welchem Ausmass Verzerrungen bei der MRA entstehen und welche Konsequenzen sich daraus ergeben. (Eine wichtige Hilfe leistet dabei die Residuenanalyse, welche Informationen über die Datenstruktur und über mögliche Verzerrungseffekte liefern kann). Wird von der Durchführung einer MRA abgesehen, muss in Erwägung gezogen werden, welche Vorteile dabei allenfalls aufgegeben werden. Zu den Vorteilen parametrischer Verfahren gehören u.a. folgende Aspekte: eine hohe Teststärke (power), die Erfassung des Standardfehlers der Messung, die zumindest teilweise Berücksichtigung von Messfehlern in den Variablen, eine gut entwickelte Messtheorie sowie umfangreiche Computerprogramme mit der Möglichkeit, grosse Datensätze mit relativ geringem Speicher- und Rechenaufwand bearbeiten zu können. Entscheidet sich der Untersucher nach dem Abwägen der Vor- und Nachteile trotz Verletzung einiger Voraussetzungen für die Durchführung einer MRA, so wird es wahrscheinlich notwendig sein, ergänzende Analysen am gleichen Datensatz durchzuführen (analog der oben dargestellten Zusatzberechnungen). Für den vorliegenden Datensatz hat die Residuenanalyse u.a. ergeben, dass die Signifikanzbestimmung der Betakoeffizienten infolge der Ver-

⁸ Die MCA zeigte zusätzlich, dass andere unabhängige Variablen, welche gemäss Original-MRA nicht bedeutungsvoll waren, relativ hohe Betakoeffizienten aufwiesen.

letzung verschiedener Annahmen nicht genügend abgesichert ist. Andererseits geht aus den Zusatzanalysen hervor, dass es die "signifikanten" Prädiktoren waren, die bei den verschiedenen Zusatzberechnungen am besten reproduzierbar waren. Im vorliegenden Datensatz wird deshalb die Signifikanzbestimmung im Sinne einer Interpretationshilfe verwendet.

Der Untersucher wird wohl auf einen konkreten Datensatz und einen theoretischen Hintergrund bezogen entscheiden müssen, ob die Durchführung einer MRA für eine bestimmte Fragestellung gerechtfertigt ist oder nicht. Dabei ist es durchaus möglich, dass trotz Verletzung von Vorschriften mit der MRA ein Beitrag zum Erkenntnisfortschritt geleistet werden kann (Allerbeck, 1978, 212). Selbstverständlich bleibt das Postulat aufrechtzuerhalten, dass die Messungen verbessert und metrisiert werden sollen und dass robuste Verfahren oder solche, welche nur Ordnungsbeziehungen verwenden, benützt werden sollen (Wainer & Thissen, 1976). Eine pauschale Kritik an der Verwendung parametrischer Verfahren in der Sozialforschung scheint uns jedoch nicht angebracht.

BIBLIOGRAPHIE

- ALLERBECK Klaus R. (1978), "Messniveau und Analyseverfahren - Das Problem 'strittiger Intervallskalen'", *Zeitschrift für Soziologie*, 7, 199-214.
- ANSCOMBE F.J. & TUKEY John W. (1963), "The Examination and Analysis of Residuals", *Technometrics*, 5, 141-160.
- BAKER Bela O., HARDYCK Curtis D. & PETRINOVICH Lewis F. (1966), "Weak Measurements vs. Strong Statistics: An Empirical Critique of S.S. Stevens' Proscriptions on Statistics", *Educational and Psychological Measurement*, 26, 291-309.
- BOHRNSTEDT George W. & CARTER Michael T. (1971), "Robustness in Regression Analysis", in COSTNER Herbert L., Ed., *Sociological Methodology*, Jossey-Bass, San Francisco, 118-146.
- BOLLEN Kenneth A. & BARB Kenney H. (1981), "Pearson's and Coarsely Categorized Measures", *American Sociological Review*, 46, 232-239.
- BRADLEY James V. (1978), "Robustness ?", *Br. J. Math. Statist. Psychol.*, 31, 144-152.
- BRADLEY James V. (1980), "Nonrobustness in Classical Tests on Means and Variances: A Large-Scale Sampling Study", *Bulletin of the Psychonomic Society*, 15, 275-278.
- BRADLEY James V. (1980), "Nonrobustness in Z, t and F Tests at Large Sample Sizes", *Bulletin of the Psychonomic Society*, 16, 333-336.
- DRAPER Norman R. & SMITH Harry (1966), *Applied Regression Analysis*, Wiley, New York.
- DURBIN James & WATSON G. (1971), "Testing for Serial Correlation in Least Squares Regression III", *Biometrika*, 58, 1-19.

- GAENSSLEN Hermann & SCHUBOE Werner (1973), Einfache und komplexe statistische Analyse, Reinhardt, München.
- HAGER Willi & WESTERMANN Rainer (1983), "Planung und Auswertung von Experimenten, in BREDEKAMP Jürgen & FEGER Hubert, Hrsg., Enzyklopädie der Psychologie, Bd. 5, Hypothesenprüfung, Hogrefe, Göttingen, 24-238.
- HARTUNG Joachim, ELPELT Bärbel & KLOESENER Karl H. (1982), Statistik. Handbuch der angewandten Statistik, Oldenburg, München.
- HORTON Raymond L. (1978), The General Linear Model. Data Analysis in the Social and Behavioral Sciences, Mc Graw-Hill, New York.
- JOHNSTON John (1972), Econometric Methods, Mc Graw-Hill, New York (2nd Ed.).
- KIM Jae O. (1976), "Multivariate Analysis of Ordinate Variables", American Journal of Sociology, 81, 261-298.
- LABOVITZ Sanford (1967), "Some Observations on Measurement and Statistics", Social Forces, 46, 151-160.
- LABOVITZ Sanford (1970), "The Assignment of Numbers to Rank Order Categories", American Sociological Review, 35, 515-524.
- LANTERMANN Ernst-D. (1976), "Zum Problem der Angemessenheit eines inferenzstatistischen Verfahrens", Psychologische Beiträge, 18, 99-104.
- LEWIS-BECK Michael S. (1980), Applied Regression. An Introduction, Sage, Beverly Hills, California.
- MAYER Lawrence S. (1970), "Comment on 'The Assignment of Numbers to Rank order Categories'", American Sociological Review, 35, 916-917.
- MAYER Lawrence S. (1971), "A Note on Treating Ordinal Data as Interval Data", American Sociological Review, 36, 519-520.
- SCHOENFELD P. (1971), Methode der Oekonometrie, Bd. II, Vahlen, München.
- SCHUBOE Werner, HAAGEN Klaus & OBERHOFER Walter (1983), "Regressions- und kanonische Analyse", in BREDEKAMP Jürgen & FEGER Hubert, Hrsg., Enzyklopädie der Psychologie, Bd. 4, Strukturierung und Reduzierung von Daten, Hogrefe, Göttingen, 207-292.
- SIEBER Martin F. & ANGST Jules (1981), Drogen-, Alkohol- und Tabakkonsum. Ein Beitrag zur Epidemiologie und Aetiologie bei jungen Erwachsenen, Huber, Bern.
- URBAN Dieter (1982), Regressionstheorie und Regressionstechnik, Teubner, Stuttgart.
- WAINER Howard & THISSEN David (1976), "Three Steps towards Robust Regression", Psychometrika, 41, 9-34.
- WOTTAWA Heinrich (1971/82), "Zum Problem der Abtestung der Verteilungsvoraussetzung in Varianz- und Regressionsanalyse", Arch. Psychol., 134, 257-263.

