Messung komplexer Variablen als Summe von Punktzahlen

Eine beliebte Methode des measurement without theory

von

Peter von der Lippe und Andreas Kladroba

Zusammenfassung

Es ist eine sehr beliebte Methode, eine "komplexe" mehrdimensionale Eigenschaft dadurch messbar zu machen, dass man "Punkte" für mehrere als "relevant" erachtete "Indikatoren" bildet und diese dann summiert (womit ex definitione eine eindimensionale Variable entsteht). Man spricht hier von Indexmethode (oder besser Punktsummenmethode, PSM). Es ist auffallend, dass die PSM einerseits sehr verbreitet ist, es aber andererseits kaum Versuche gibt, sich mit ihren inhaltlichen und formalen Grundlagen auseinanderzusetzen. Es gibt nicht nur keine wirkliche Begründung für die Summenbildung, es ist sogar schon unklar, ob die Indikatoren korreliert oder unkorreliert sein sollten. Der Beitrag stellt Zusammenhänge mit multivariaten Verfahren (Faktorenanalyse, Clusteranalyse) dar und problematisiert die Rolle der Gewichtung.

Summary

Measuring complex variables by sums of scores

A popular and widespread, though rarely ever scrutinised method of measuring a complex multidimensional construct y consists of adding the scores $x_1, ..., x_m$ assigned to a unit in m "dimensions". These scores ("indicators") are supposed to represent the m relevant aspects of y. Interestingly the sum-of-scores method (SSM) is not only lacking a theoretical justification there is not even a clear understanding of the relevance of the correlation between the indicators for y. The paper therefore aims at relating the SSM to methods of multivariate analysis and tries to clarify the role of weights in the summation of scores.

JEL - Klassifikation: E3, C0, I3, M0

1. Gegenstand der Punktsummenmethode (PSM)

a) Warum ist die Methode so beliebt?

Gegenstand dieses Beitrags ist eine Methode, die in der empirischen Sozialforschung bekannt ist unter dem Namen "Indexmethode", die wir jedoch, um Verwechslungen zu vermeiden "Punktsummenmethode" (PSM) nennen wollen. Sie besteht einfach darin, eine als "komplex" geltende (viele "Dimensionen" umfassende) Eigenschaft[1] dadurch meßbar zu machen, dass

[1] Beispiele für derartige komplexe Variable sind "Konjunktur", Wettbewerbsfähigkeit, Liberalität (eines Landes) oder "Lebensqualität" (Wohlfahrt). Die verschiedensten Arten von "rankings" beruhen auf "Indizes" in diesem Sinne. Auch bei neuerdings in Mode gekommenen Versuchen der "multidimensionalen Messung der Armut", wird gnadenlos summiert (auch ohne Gewichtung) über Indikatoren wie Nettoeinkommen, Haushaltsgröße, "residence and tenure status", Wohnungsqualität (Ausstattung mit Bad usw.), Umweltvariablen, Indikatoren der Kriminalität usw. In allen diesen Fällen wird letztlich eine Messung durch eine Punktsumme versucht. Ein interessantes neueres Beispiel aus der empirischen Konjunkturanalyse mit der PSM ist der "index of leading indicators at EU level", vgl. S. Pattanaik, Eurostat is half way there, in: Sigma 2/2002, S. 11ff.

man eine gewogene (z*) oder ungewogene Summe (z)[2] von "Punkten" für m > 1 jeweils als "relevant" erachtete "Indikatoren" oder "Einzelvariablen" (EVn) bildet, für die Meßwerte x_i, i = 1, ..., m auf einer metrischen Skala postuliert, bzw. irgendwie konstruiert werden

(1) $z = x_1 + x_2 + ... + x_m$,

(2) $z^* = b_1x_1 + b_2x_2 + ... + b_mx_m$.

Dabei wird in der Regel $\sum b_i = 1$ angenommen, so dass z* ein arithmetisches Mittel ist ($z^* = \bar{x}$). Die Summenbildung mit oder ohne Gewichtung der Summanden ist die einfachste Art, aus einer Vielzahl metrisch skalierter Größen eine einzige Größe zu machen, also eine eindimensionale Messung[3] zu erreichen, die wohl deshalb so sehr gewünscht wird, weil es nur dann möglich ist, eine durchgängige Rangordnung der Einheiten aufzustellen. Das Verfahren ist denkbar einfach und (wohl auch deshalb) sehr beliebt, andererseits aber (erstaunlicherweise) wenig untersucht worden. Man könnte es als eine Primitivmethode der multivariaten Analyse bezeichnen; es ist ein "klassischer Fall" von "measurement without theory". Es ist eine Methode "deren einzige Begründung in der Meßvorschrift selbst besteht, d.h. in der Art und Weise wie den einzelnen Objekten Zahlen zugeordnet werden"[4].

Eine komplexe Variable durch eine Punktsumme messen zu wollen verlangt drei Schritte

1. Auswahl vom m "Indikatoren" (Komponenten oder "Einzelvariablen", EVn)
2. Skalierung der EVn und Sammlung der Daten für die EVn[5]
3. Einführung eines Wägungsschemas und Berechnung der gewogenen Summe[6].

Die Einfachheit dieser Vorgehensweise mag dazu verleiten, alles und jedes mit dieser "Methode", also als eine Summe z oder z* messen zu wollen.

b) Rechtfertigung der Methode

Die Punktsummenmethode (PSM) läßt sich kaum rechtfertigen. Das gilt heutzutage umso mehr, wo es leicht möglich ist, auch sehr aufwändige Rechenarbeiten durchzuführen und die Software für kompliziertere multivariate Verfahren, wie z.B. die Faktorenanalyse jedermann zugänglich ist. Nach unserer Kenntnis gibt es im Wesentlichen nur die folgenden vier Versuche, der Methode einen Sinn zu geben, die schwerlich alle gemeinsam zutreffend sein können:

1. Jede EV ist für sich genommen ein fehlerbehaftetes Maß der Variable, die es eigentlich zu messen gilt. Bei mehreren EVn erhöht sich somit "die *Chance, Meßungenauigkeiten zu verringern* und so den gemeinten 'wahren' Sachverhalt eher abzubilden"[7]. Die Zusammen-

[2] In der Notation soll im Folgenden unterschieden werden zwischen einer Punktsumme z = Σx und der eigentlich intendierten komplexen Variable y.
[3] Eine Messung ist eindimensional (mehrdimensional), wenn sich das Ergebnis durch eine Zahl, ein Skalar (oder mehrere Zahlen, einen Vektor) ausdrücken läßt. Eine mehrdimensionale Messung wird vom Benutzer der Statistik zu recht als eine Leistung minderen Ranges empfunden, und sie gilt als schwerer interpretierbar.
[4] H. Kromrey, Empirische Sozialforschung, 5. Aufl., Opladen 1991, S. 180.
[5] Hier wird oft ein gewaltiger Aufwand mit der Sammlung von Daten betrieben, wobei diese dann in der anschließenden Auswertung oft durch Nichtbeherrschung der ohnehin schon sehr simplen statistisch-methodischen Grundlagen der "Indexmethode" zu einem ziemlich wertlosen Zahlenfriedhof gemacht wird. Ein Beispiel hierfür ist der Atlas der Lebensbedingungen in Deutschland von D. Korczak, Lebensqualität-Atlas, Opladen, 1995. Wir gehen im folgenden auf diesen Atlas gelegentlich ein; vgl. auch P. von der Lippe, Statistische Wohlfahrtsindikatoren - Die Messung des Lebensstandards, in: Statistisches Bundesamt (Hrsg.), Wohlfahrtsmessung, Aufgabe der Statistik im gesellschaftlichen Wandel, Band 29 der Schriftenreihe Forum der Bundesstatistik, Stuttgart 1996, S. 39 - 72 mit weiteren Literaturhinweisen.
[6] Neben der Addition wurde auch gelegentlich eine Multiplikation (also ein geometrisches Mittel) vorgeschlagen.
[7] Kromrey, a. a. O., S. 121.

fassung mehrerer EVn dient somit der "*Messfehlerreduktion*"[8]. Hier wird eine Überlegung, die unbestritten ist für die Schätzung eines arithmetischen Mittels μ mit $\hat{\mu} = \bar{x}$ aus mehreren fehlerbehafteten Einzelwerten $x_1 = \mu + u_1$, $x_2 = \mu + u_2$ usw. *der gleichen einen* (identisch verteilten) Zufallsvariable X mit den *zufälligen* Meßfehlern u_1, u_2, ... übertragen auf die Schätzung einer Variable aus *mehreren* (unterschiedlichen) Variablen. Betrachtet man Anwendungen der PSM, etwa die Messung des sozialen Status mit EVn wie Einkommen (E), Beruf (B) und Schulbildung (S), so fällt es schwer sich vorzustellen, dass E, B und S im Grunde alle das gleiche darstellen, gäbe es nicht die Störung durch einen Meßfehler. In diesem Sinne wird auch gesagt, die Summation über die EVn beseitige die *Redundanz*[9], die dadurch gegeben ist, dass die EVn in einer nicht näher bekannten Weise alle das gleiche messen.

Es ist nicht klar, ob sich aus dieser Vorstellung ableitet, dass die EVn

- untereinander alle positiv korreliert sein sollten, einfach weil sie das gleiche messen und es deshalb auch besser wäre, weniger statt mehr *redundante* EVn zu verwenden
- oder ob nicht umgekehrt die Messung umso besser (im Sinne des Fehlerausgleichs) ist, je mehr EVn verwendet werden[10], die dann auch negativ korreliert sein sollten.

Mit dieser Argumentation ist es auch schwer in Einklang zu bringen, warum die EVn gewichtet werden sollten.

2. Nicht selten wird *auch zusammen mit der Rechtfertigung Nr. 1* (!) das folgende Argument[11] angeboten: jeder Indikator bildet nur einen "Teil" des eigentlich zu messenden Konzepts ab. Die Indikatoren messen also gerade *nicht* das gleiche. Aber es fehlt hier jede Erklärung dafür

   - wie viele EVn benötigt werden und woran man erkennt, ob es zu wenige oder zu viele sind
   - ob die EVn positiv oder negativ[12] korreliert oder am besten unkorreliert sein sollten,
   - ob (und wenn ja wie) ein Zusammenhang zwischen der Größe des "Teils" und der Größe des Gewichts der EV besteht.

3. Eine noch mehr im Vagen bleibende Vorstellung ist die Idee des "universe of attributes", womit eine *Auswahl von Variablen* zur Erfassung der zu messenden Eigenschaft in ihrer wirklichen Komplexität *nach Art einer Stichprobe* aus der Gesamtheit möglicher Variablen gerechtfertigt erscheint. Es bleibt aber offen, welche Beziehungen zwischen den Variablen gelten und wie eine Summierung zu rechtfertigen ist[13], ganz abgesehen davon, dass - wie

[8] ebenda, S. 121.
[9] Auch zu dieser Interpretation vgl. Kromrey, S. 121.
[10] Man beachte, dass man demgegenüber vom Standpunkt der Faktoren- und Clusteranalyse eher geneigt ist, wenige statt viele EVn als Vorteil zu empfinden.
[11] "Jeder Indikator ist für sich genommen nur eine teilweise (operationale) Definition eines Begriffs", Kromrey, a. a. O., S. 121.
[12] Für eine negative Korrelation spricht, dass bei zwei Teilen y_1 und y_2 des "gesamten" Konzepts $y = y_1 + y_2 =$ const y_1 umso größer (kleiner) ist, je kleiner (größer) y_2 ist. Außerdem spricht dafür dass bei einer Punkt*summe* implizit unterstellt wird, dass "sich die Werte auf den als Indikatoren herangezogenen Variablen gegenseitig kompensieren (dass also ein niedriger Wert auf Indikator 1 durch einen hohen Wert auf Indikator 2 ausgeglichen werden kann)", ein Gedanke; der sich auch in der negativen Steigung der Geraden gleicher Gesamtpunktzahl in Abb. 2 widerspiegelt. In Abschn. 3 dieses Beitrags zeigen wir dass man bei *negativer* Korrelation zwischen den EVn in Schwierigkeiten kommen dürfte.
[13] Es gibt Fälle, in denen eine Variable x selbst eine Summe darstellt, etwa das Gesamteinkommen als Summe von x_1, etwa dem Einkommen aus unselbständiger Arbeit, x_2 dem Einkommen aus Vermietung und Verpachtung

noch gezeigt wird - die Analogie zwischen einer Auswahl von Einheiten und einer Auswahl von Variablen wenig durchdacht zu sein scheint.

4. Beziehungen zwischen den EVn und der Zielgröße y werden explizit ins Spiel gebracht bei Versuchen, die Indexmethode mit Konzepten der Faktorenanalyse (FA) zu rechtfertigen[14]. Die mit der FA bestimmten "Faktoren" sind ebenfalls Linearkombinationen der EVn[15], was nicht nur die Bildung von Punktsummen zu rechtfertigen scheint, sondern auch die oft vorgenommene Bestimmung von Gewichten aufgrund der Ladungen im ersten Faktor. Gelegentlich finden sich auch einige Versuche, mit der FA verwandte Methoden heranzuziehen, wie z.B. die Latent-Structure Analysis (LSA)[16], oder die hinsichtlich der unterstellten Kausalstruktur flexiblere (als die FA) Pfadanalyse[17] bzw. deren Weiterentwicklung LISREL.

Auf die ersten beiden Rechtfertigungsversuche sollte kurz eingegangen werden[18] Angenommen es gelte im Sinne des ersten Arguments in der Tat für jede EV x_{ij} (bei i = 1, ..., m EVn und j = 1, ..., n Beobachtungseinheiten) und der wahren (intendierten) Größe y_j der folgende Zusammenhang $x_{ij} = y_j + u_{ij}$. Welchen Sinn hat dann beispielsweise die Bildung eines gewogenen Mittels von der Einfachheit halber nur m = 2 Indikatoren? Es gilt dann

$$z_j = b_1 x_{1j} + b_2 x_{2j} = (b_1 + b_2) y_j + (b_1 u_{1j} + b_2 u_{2j}) = y_j + v_j .$$

Wenn wie üblich $0 \leq b_1, b_2 \leq 1, b_1 + b_2 = 1$ angenommen wird, kann die Varianz von v, nämlich $\sigma_v^2 = b_1^2 \sigma_{u_1}^2 + b_2^2 \sigma_{u_2}^2 + 2 b_1 b_2 \sigma_{u_1 u_2}$, in der Tat kleiner sein kann als die von u_1 und u_2 jeweils für sich genommen[19], also $\sigma_{u_1}^2$ und $\sigma_{u_2}^2$. So gilt etwa im folgenden Fall $b_1 = 0{,}3$, $\sigma_{u_1}^2 = 4$ und $\sigma_{u_2}^2 = 3$, wenn die Kovarianz $\sigma_{u_1 u_2} = -2$ beträgt (so dass u_1 und u_2 mit -0,5773 korrelieren): $\sigma_v^2 = 0{,}36 + 1{,}47 - 0{,}84 = 1{,}83 - 0{,}84 = 0{,}99$. Gilt aber $\sigma_{u_1 u_2} = +\,3{,}4$ (Korrelation $r_{u_1 u_2} = 0{,}9815$), so erhält man für σ_v^2 den Wert $1{,}83 + 1{,}428 = 3{,}258 > \sigma_{u_2}^2 = 3$. Das bedeutet, dass man besser gefahren wäre, wenn man die unbekannte Variable y durch x_2 statt durch eine gewogene Summe von x_1 und x_2 gemessen hätte. Würde man das größere Gewicht 0,7 nicht der weniger streuenden Variable x_2 sondern der Variable x_1 geben, dann wäre σ_v^2 sogar 2,23 + 1,428 = 3,658 > 3,258. Soll das Argument der Messfehlerreduktion gelten, so wäre es vorteilhaft, wenn die Störgrößen (Meßfehler) $u_1, ..., u_m$ negativ korrelieren würden. Andererseits wäre es dann jedoch schwer geltend zu machen, dass alle x-Variablen $x_1, ..., x_m$ letztlich das gleiche messen.

---

usw. Aber es wäre unsinnig, dann verschiedene Stichproben als dem "universe" mit den Variablen $x_1, x_2, x_3, ...$, etwa vom Umfang n = 3, zu akzeptieren, also etwa $x_1 + x_2 + x_3$ ebenso wie $x_1 + x_3 + x_5$ oder $x_2 + x_5 + x_7$ usw. als Schätzer für das Gesamteinkommen x. Eine Auswahl zulassen und eine Summation fordern verträgt sich schlecht miteinander.

[14] Zu dieser Argumentation vgl. K. Holm, Theorie der Fragenbatterie, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Bd. 26 (1974), S. 316-341.

[15] Ebenda, S. 341.

[16] Eine Teilgesamtheit befindet sich auf dem gleichen Niveau y_0 einer latenten Variable y und ist in diesem Sinne "homogen", wenn je zwei manifeste Variablen x_i und x_j in ihr nicht korrelieren ("lokale Unabhängigkeit" bei der LSA, bzw. die partielle Korrelation verschwindet $r_{ij.y} = 0$). Im Detail berufen sich auf die LSA zur Rechtfertigung der Indexmethode Besozzi und Zehnpfennig, a. a. O., und (kritisch der Indexmethode gegenüber) Wärneryd, B., Some Aspects of Index Formation, in: Acta Sociologica 7 (1963), S. 19-32.

[17] Über Versuche, auf diese Weise Skalen für latente Variablen zu konstruieren vgl. B. Wegener: Gibt es Sozialprestige?, in: Zeitschrift für Soziologie 1985, Bd. 14/3, S. 209-235.

[18] Wir halten sie für ungeeignet weil sie von nicht genügend durchdachten Analogien ausgehen. Auf die dritte Gruppe von "Argumenten" soll wegen der Vagheit der sie tragenden Konzepte nicht weiter eingegangen werden.

[19] Es ist unschwer zu erkennen, dass die folgende Betrachtung ähnlich der ist, die man von der Rechtfertigung der Risikominderung durch Risikostreuung in der Theorie des portfolio selection her kennt.

Aus dem zweiten Rechtfertigungsversuch könnte sich der Gedanke ableiten, dass etwa bei m = 2 gilt $x_{ij} = y_{ij} + u_{ij}$, i = 1, 2, so dass $x_{1j} + x_{2j} = y_{1j} + y_{2j} + u_{1j} + u_{2j}$ (man stelle sich also einmal im einfachsten Fall das Bild von "Teilen" eines Konstrukts y als Summanden die sich zu y summieren vor). Eine solche Betrachtung würde Sinn machen, wenn der "wahre Wert" der Einheit j auf der zu messenden komplexen Variable y in der Tat die Summe von y_1 und y_2 darstellt, also $y_j = y_{1j} + y_{2j}$. Unter diesen Voraussetzungen

- macht eine Gewichtung wenig Sinn (y ist die *un*gewogene Summe unabhängig davon, ob y_1 oder y_2 der größere "Teil" von y ist), und
- es sollten die EVn x_1 und x_2 korreliert sein[20], wie im Folgenden aus Gl. 6 hervorgeht und bei[21] $E(u_1) = E(u_2) = E(y_1u_1) = E(y_2u_2) = 0$ und deshalb $E(x_i) = E(y_i) = \mu_i$ (i = 1, 2) die folgenden Beziehungen gelten

(3) $\sigma_{x_1x_2} = E(x_1x_2) - E(x_1)E(x_2) = E(y_1y_2) - E(y_1)E(y_2) + E(u_1y_2) + E(u_2y_1) + E(u_1u_2)$ oder

(3a) $\sigma_{x_1x_2} = \sigma_{y_1y_2} + E[(u_1 + u_2)y] + E(u_1u_2)$

und ganz analog erhält man

$$\sigma_{x_1y} = E(x_1y) - E(x_1)E(y) = E[(y_1 + u_1)y] - \mu_1^2 - \mu_1\mu_2 = E(y_1^2) + E(y_1y_2) + E(u_1y) - \mu_1^2 - \mu_1\mu_2$$

oder

(4) $\sigma_{x_1y} = \sigma_{y_1}^2 + \sigma_{y_1y_2} + E(u_1y)$ und entsprechend

(5) $\sigma_{x_2y} = \sigma_{y_2}^2 + \sigma_{y_1y_2} + E(u_2y)$, und wegen

$\sigma_y^2 = \sigma_{y_1}^2 + \sigma_{y_1y_2} + \sigma_{y_2}^2 + \sigma_{y_1y_2} = \sigma_{x_1y} + \sigma_{x_2y} - E[(u_1 + u_2)y]$ und $\sigma_{x_i}^2 = \sigma_{y_i}^2 + \sigma_{u_i}^2$ ist auch

(3b) $\sigma_{x_1x_2} = \sigma_{x_1y} + \sigma_{x_2y} - \sigma_{y_1y_2} - \left(\sigma_{y_1}^2 + \sigma_{y_2}^2 - E(u_1u_2)\right)$

(6) $$r_{x_1x_2} = \frac{\sigma_{x_1x_2}}{\sigma_{x_1}\sigma_{x_2}} = r_{x_1y}\frac{\sigma_y}{\sigma_{x_2}} + r_{x_2y}\frac{\sigma_y}{\sigma_{x_1}} - r_{y_1y_2}\frac{\sigma_{y_1}\sigma_{y_2}}{\sigma_{x_1}\sigma_{x_2}} - \frac{R}{\sigma_{x_1}\sigma_{x_2}}$$

mit $R = \sigma_{y_1}^2 + \sigma_{y_2}^2 - E(u_1u_2)$. Man beachte, dass sich somit *nicht* eine Beziehung nach Art der Faktorenanalyse (bei *einem* Faktor y, dem "Generalfaktor"[22]) ergibt, wonach zu erwarten wäre dass gilt $r_{x_1x_2} = r_{x_1y}r_{x_2y}$. Aus Gl. 3b erhält man schließlich in Verbindung mit Gl. 4 und 5 auch für die Kovarianz $\sigma_{x_1x_2}$

(7) $\sigma_{x_1x_2} = \sigma_{y_1y_2} + E(u_1u_2) + E[(u_1 + u_2)y]$

Man beachte, dass nach Gl. 4 und 5 die Kovarianz σ_{x_1y} (und entsprechend σ_{x_2y}) nicht abhängig ist von der Kovarianz zwischen den EVn also $\sigma_{x_1x_2}$, wie dies bei der Punktsumme z im Unterschied zu y der Fall ist (vgl. Gl. 9), sondern hierfür (im Falle von σ_{x_1y}) die Varianz von

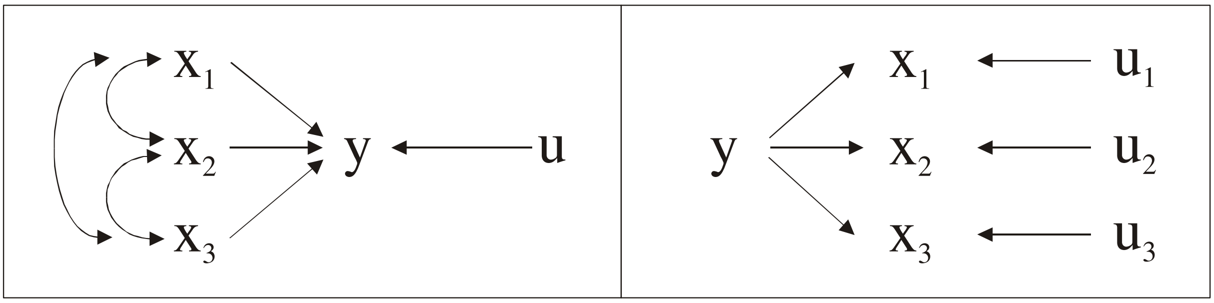
[20] Kromrey, S. 181f folgert aus dem Gedanken, dass jede Variable nur einen Teil des zu messenden Konstrukts abdeckt, dass dies nur möglich sei, wenn die Indikatoren "unabhängig" (!) seien. Andererseits findet man bei ihm aber auch im Zusammenhang mit einem Versuch, die PSM mit dem Modell der multiplen Regression (Abb. 1a) zu rechtfertigen, die Vorstellung, dass alle Indikatoren entweder positiv oder negativ mit der (im Fall der PSM allerdings nicht vorhandenen) unabhängigen Variable y korreliert sein sollten (a. a. O., S. 121).

[21] Im folgenden bleiben wir bei der Schreibweise mit kleinen Buchstaben für die (Zufalls-)Variablen.

[22] vgl. Abschn. 2b.

y_1 (wofür x_1 als Indikator dient) und die Kovarianz $\sigma_{y_1y_2}$ entscheidend ist (analog $\sigma^2_{y_2}$ und $\sigma_{y_1y_2}$ im Falle von σ_{x_2y}).

Abb. 1a Abb. 1b



Die vierte Gruppe von Rechtfertigungsversuchen betrifft in der Tat Methoden, auf die auch im folgenden Bezug genommen wird. Solche Methoden *können* brauchbar gemacht werden für eine Konstruktion oder Überprüfung von "Indizes" z im Sinne der PSM. Dabei ist aber zu beachten, dass es ein erheblicher Unterschied ist, ob sich mit einer FA tatsächlich eine latente Dimension in den Daten feststellen ließ und sich hierauf eine Aussage stützt, oder ob man nur das in der Praxis übliche Vorgehen relativ *willkürlich* bestimmte Variablen als "relevant" heraus zu greifen, sie zu skalieren und zu gewichten mit der FA glaubt theoretisch rechtfertigen zu können.

Auch hier besteht eine große Unklarheit. So wird z.B. die PSM mit einem Kausalmodell nach Art der Abb. 1a gerechtfertigt[23], ein Modell dass dem der multiplen Regression entspricht in dem die *beobachtete* Variable y mit drei Regressoren x_1, x_2 und x_3 erklärt wird. Angemessener dürfte jedoch das Modell der Abb. 1b sein, das der Faktorenanalyse mit einem Generalfaktor y (der keine beobachtete Variable ist) zugrunde liegt.

In *beiden* Fällen (Abb. 1a und 1b) ist der Wert y_j oder $\hat{y}_j$ der Einheit j bezüglich der komplexen Variable y eine (gewogene) Linearkombination der EVn-Werte, also x_{1j}, x_{2j} und x_{3j}. Dass bei der PSM eine solche Linearkombination gebildet wird, ist allein also noch kein Hinweis auf eine Verwandtschaft dieser Methode mit der Regressions- oder Faktorenanalyse.

Im Ergebnis läßt sich festhalten, dass keine schlüssige oder gar sauber statistisch ausformulierte Modellvorstellung für die PSM existiert, aus der die Berechtigung der einzelnen Verfahrensschritte hervorginge. Das ist umso bedauerlicher als ziemlich allgemein bekannt ist, dass das praktische Vorgehen unbefriedigend ist (Abschn. 1c) und es auch an einem externen Kriterium der Validierung fehlt (Abschn. 1d).

c) Inhaltliche Kritik

Es fällt auf, dass in der Öffentlichkeit und in der Literatur in der Regel eher nur die inhaltlichen Aspekte der PSM kritisiert werden, wie z.B. die Auswahl der Indikatoren oder die Gewichtung, aber kaum formale Aspekte oder die generelle Berechtigung der Methode.

Die Summe z* macht den *Eindruck* einer metrischen Skala für eine Dimension. Dass dies nur oberflächlich betrachtet ein Eindruck ist, wird deutlich, wenn man sich fragt, was denn die Maßeinheit dieser Dimension sein soll.[24] Weil eine Punktzahl keine Entsprechung in der Rea-

[23] vgl. Kromrey a. a. O., S. 121. Die Notation in Abb. 1a ist im Vergleich zu Kromrey etwas geändert. Doppelpfeile bedeuten, dass die Regressoren x_1, x_2 und x_3 untereinander korreliert sein dürfen, obgleich sie nicht in einer kausalen Beziehung zueinander stehen.

[24] Es gibt keine Antwort auf die naheliegende Frage welche Maßeinheit bei einer Summe von Punktzahlen vorliegt.

lität hat, kann man sich auch die Freiheit nehmen, die verschiedensten Indikatoren zu einer Summe zusammenzufassen, ohne Rücksicht auf den Charakter der Variablen z.B. Bestands- und Stromgrößen (Ausstattung mit Krankenhäusern und Anzahl der Verkehrsunfälle) oder auf mögliche Doppelzählung ("Redundanzen"). Schon das ist ein Indiz für die Willkür der Methode. Ein weiteres Indiz dafür ist, dass auch das Hinzukommen weiterer Summanden in der Punktsumme als unschädlich gilt, wenn es nicht sogar als nützlich empfunden wird.

Es soll nun daran erinnert werden, dass es in der Statistik nicht sehr naheliegend ist, über Merkmale (statt über Einheiten) eine Summe zu bilden. Es ist unstrittig, dass es Sinn macht, über *Personen* bezüglich des gleichen Merkmals zu addieren, d.h. Häufigkeiten *der gleichen* Merkmalsausprägung festzustellen um so dann z.B. ein arithmetisches Mittel bezüglich des betreffenden Merkmals zu berechnen. Aber mit welcher Begründung soll eine Addition über Merkmale einer Einheit genauso zulässig sein, wie eine Addition über Einheiten bezüglich eines Merkmals?

Es gibt hierzu den bereits erwähnten Gedanken eines "universe of attributes", aus dem man einzelne Merkmale auswählen kann, so wie man aus einer Grundgesamtheit von Einheiten eine Stichprobe zieht. Aber es ist nicht gelungen, auf dieser von L. Guttman stammenden Überlegung eine befriedigende "Theorie" der PSM aufzubauen[25]. Die Auswahl von Merkmalen ist offenbar etwas grundlegend anderes als die von Personen (Einheiten)[26].

Wir referieren schließlich noch einige - etwas näherliegende und deshalb auch häufiger vorgebrachte - kritische Bemerkungen nach Maßgabe der oben genannten drei Verfahrensschritte der PSM (oder "Indexmethode"), der Auswahl, Skalierung und Gewichtung von Indikatoren:

1. Es ist üblich, dass die *Auswahl von Indikatoren* (items) bei der PSM "der ungebremsten Phantasie der Skalenkonstrukteure" entspringt[27], so dass man auch von einer "willkürlichen Messung" sprechen kann[28]. Mehr als eine vage Vorstellung darüber, dass diese oder jene Erscheinung, für das, was zu messen ist, "wichtig"[29] ist, kann der Indexkonstrukteur in der Regel nicht aufbieten. Dies ist der Verfahrensschritt, bei dem die meiste Phantasie gefordert ist und bei dem (deshalb?) auch die Kritik besonders gerne ansetzt, weil man glaubt, dass einzelne Komponenten (EVn) fehlen, oder überflüssig seien oder ihnen nicht das richtige Gewicht verliehen worden sei. Nicht selten ist es auch so, dass die EVn schon begrifflich schwer voneinander abzugrenzen sind (geschweige denn ihre Auswahl mehr als nur intuitiv begründet werden kann). So wurden kürzlich in einem Versuch, die "Qualität" von US-Präsidenten zu messen Professoren der Geschichte und Politikwissenschaft gefragt nach einer Bewertung von Präsidenten auf einer Fünf-Punkte-Skala hinsichtlich 20 EVn worunter u.a. Dimensionen wie overall-, executive- und leadership-ability fielen[30].

2. Bei der Skalierung der EVn geht es vor allem darum, die EVn hinsichtlich ihres Wertebereichs kommensurabel zu machen und zu verhindern, dass die eine oder andere EV al-

---

[25] Vgl. C. Besozzi und H. Zehnpfennig: Methodologische Probleme der Indexbildung, in: J. van Koolwijk und M. Wieken-Mayser, Techniken der empirischen Sozialforschung, Bd. 5, Testen und Messen, München, Wien 1976, S. 9ff. (32ff).

[26] Es ist leicht zu sehen, dass die Vergabe von Punkten (die keine Maßeinheit haben) u.a. den Zweck erfüllt, diese Unterschiedlichkeit zu verwischen.

[27] Vgl. R. Schnell, P.B. Hill und E. Esser, Methoden der empirischen Sozialforschung, 3. Aufl., München, Wien 1992, S. 193.

[28] ebenda, S. 183.

[29] Vgl. Besozzi und Zehnpfennig a.a.O., S. 47 für einen nicht sehr weit führenden Versuch, dieses Konzept im Rahmen der PSM zu operationalisieren.

[30] D. Lonnstrom, U.S. Presidents: A Statistical Rating, in: Bulletin of the International Statistical Institute, 54th Session, Berlin 13 - 20 August 2003, Contributed Papers, Vol. LX, Book 1, p. 745.

lein schon wegen ihrer Standardabweichung besonders stark "ins Gewicht" fällt[31]. So wurde z.B. in dem erwähnten Lebensqualitäts-Atlas der Bereich (materieller) "Wohlstand" mit so unterschiedlichen (schon hinsichtlich ihrer Maßeinheit und Größenordnung) EVn "gemessen" wie Lohn- und Gehaltssumme, Arbeitslosenquote (x_1), Anzahl der Sozialhilfeempfänger und Baulandpreise, oder der Bereich "Gesundheit" mit der Todesrate, der Zahl der Kreislauf- und Atemwegserkrankungen (x_2) usw. Es wäre unsinnig z.B. die Zahl Kreislauf- und Atemwegserkrankungen und die Arbeitslosenquote in ihrer natürlichen Maßeinheit zu mitteln ohne vorher eine entsprechende Transformation zur Variable x_1 und x_2 durchgeführt zu haben. Auch bei dieser Transformation gibt es keine inhaltlich motivierten Hilfen, was wohl genauso auch für das folgende Problem der Gewichtung gilt.

3. Bei der Gewichtung der EVn soll erreicht werden, dass die einzelnen "Indikatoren" das Ergebnis nach Maßgabe ihrer "Wichtigkeit" beeinflussen. Nur Indikatoren zu haben und diese gewichten zu wollen, das sind Probleme, die bei einer Wertsumme (etwa beim BSP) nicht auftreten. Auch die Gewichtung bei einem Preisindex ist von anderer Art, weil eine gewogene Summe von Preismesszahlen als Ausgabenvergleich interpretiert werden kann, eine gewogene Summe von Punktzahlen aber nicht. Neben dieser expliziten gibt es auch eine implizite Gewichtung. Eine implizite Gewichtung einzelner Variablen kann nicht nur durch eine ungeeignete Skalierung der EVn entstehen, sondern auch dadurch, dass ein bestimmter Bereich durch mehr, ein anderer durch weniger EVn repräsentiert wird[32].

Bei allen diesen Verfahrensschritten sind die gefundenen "Lösungen" (oder besser: die übliche Praxis) einfach und letztlich ziemlich unbefriedigend.

d) Was tun wenn es kein externes Kriterium der Validität gibt?

Häufig wird versucht, eine Messung im Sinne einer Punktsumme z durch Korrelation mit einem externen Kriterium y zu "validieren" und bei Erfolg (hoher Korrelation zwischen z und y) gilt auf diese Weise die zugrundeliegende Meßmethode als gerechtfertigt. Das Problem ist aber, dass wir y, das "wahre" Konstrukt, also z. B. den "wahren" Wohlstand oder die "wahre" Lebensqualität nicht kennen und y auch *nicht* begrifflich gleichzusetzen ist mit der Punktsumme z. Auch in diesem Punkte steht es also schlecht um die PSM, die ja typischerweise gerade dann angewendet wird, wenn "man nicht genau weiß, was man eigentlich mißt"[33], denn könnte man das beabsichtigte Konstrukt einfach mit einer *gegebenen* Variable korrelieren, denn sonst würde man ja diese als Maß für das, was man messen möchte akzeptieren und nicht nach einer Vielzahl von Indikatoren suchen[34]. Es gibt somit kein Kriterium dafür, dass die Auswahl der Indikatoren (Summanden) für z, die Punktsumme *vollständig* und *sinnvoll* (z.B. im Sinne einer hohen positiven Korrelation mit y) ist.

---

[31] In der gleichen Weise wird i.d.R. allenfalls der standardisierte Regressionskoeffizient (β-Koeffizient) in einer Regressionsfunktion als Maß der relativen Bedeutung eines Regressors benutzt, weil hier die Unterschiedlichkeit der Standardabweichungen der Regressoren ausgeschaltet ist.

[32] In dem erwähnten Lebensqualitäts-Atlas wurde z.B. der Bereich "Umwelt" durch neun gleichgewichtete EVn repräsentiert, der Bereich "Gesundheit" dagegen nur durch drei, weshalb letzterer offenbar implizit viel weniger ins Gewicht fällt (oder auch fallen sollte), ganz abgesehen davon, dass evtl. mit EVn wie Schadstoffbelastung mit SO_2, NO_2, Ozonwerten und Schwebestaub-Immissionen einerseits und Zahl der Kreislauf- und Atemwegserkrankungen andererseits durchaus etwas ähnliches gemessen worden sein kann. Wir sind solchen Fragen mit der Faktorenanalyse nachgegangen, einer Methode, die wohl am ehesten geeignet sein dürfte, der PSM eine theoretische Fundierung zu geben.

[33] Besozzi und Zehnpfennig, a.a.O., S. 9.

[34] Aus diesem Grunde wird meist keine Chance für eine Validierung mit einem Außenkriterium gesehen, vgl. auch Besozzi und Zehnpfennig, a. a. O., S. 43, Kromrey, a. a. O., S. 182.

Die im Folgenden angestellten Überlegungen zu dem bei der PSM implizit verwendeten Konzept der "Dimension" sollen demgegenüber zeigen, dass man durchaus einer "Messung" durch eine Punktsumme etwas von ihrer Beliebigkeit nehmen kann[35].

2. Konzept der Eindimensionalität

a) Eigenschaften einer Summe als zusammenfassende komplexe Variable

Eine Punktsummen-Rechnung impliziert, dass die Punktsumme z oder z* nicht sinkt, wenn eine ihrer Komponenten (EVn) zunimmt, alle EVn sind vom Typ "je mehr, desto besser", es gibt keinen Überdruß, von dem ab eine weitere Steigerung eine Verschlechterung darstellt, die EVn sind beliebig substituierbar und additiv (es gibt keine "interactions"). Der "Grenznutzen", d.h. der Beitrag, den die Steigung einer Komponente zum Wohlstand leistet $\frac{\partial z^*}{\partial x_i} = b_i$ ist nicht nur konstant, sondern auch unabhängig davon, welchen Wert eine andere Komponente annimmt. Dabei kann man sich durchaus vorstellen, dass z.B. mehr Freizeit x_2 um so mehr Genuß darstellt, je höher das verfügbare Einkommen x_1 ist, je mehr man auch in dieser Zeit unternehmen und sich leisten kann, so dass auch ein "Index" $z^* = b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + \ldots$ mit einem *interaction term* b_{12} sinnvoll sein könnte, in dem nämlich der isolierte "Beitrag" von x_2 zu z* mit $\frac{\partial z^*}{\partial x_2} = b_{12} x_1$ von x_1 abhängen würde. Ausgeschossen sind in der üblichen Vorgehensweise bei der PSM auch Nichtlinearitäten nach Art von $z^* = b_1 x_1 + b_2 x_2^2 + b_3 \sqrt{x_3} + \ldots$ o.ä., obgleich auch diese bei bestimmten komplexen Erscheinungen z* sinnvoll sein könnten.

Im Falle des linearen Modells der Gl. 2 gilt für die Varianz der Variable z* bekanntlich (der ungewogene Fall z ist ein Spezialfall mit $b_1 = \ldots = b_m = 1$ bzw. 1/m)

$$\sigma_{z^*}^2 = b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + b_3^2\sigma_3^2 + \ldots + 2b_1b_2\sigma_{12} + 2b_1b_3\sigma_{13} + 2b_2b_3\sigma_{23} + \ldots \tag{8}$$

wobei σ_i^2 die Varianzen der Indikatoren x_i und σ_{ij} die Kovarianzen zwischen den EVn sind. Für die Kovarianz zwischen einer EV etwa x_1 und z* erhält man dann

$$\sigma_{z^*x_1} = b_1\sigma_1^2 + b_2\sigma_{12} + b_3\sigma_{13} + b_4\sigma_{14} + \ldots \tag{9}$$

so dass für die Steigung in der Regression von z* auf x_1 gilt

$\frac{\sigma_{z^*x_1}}{\sigma_1^2} = b_1 + b_2\frac{\sigma_2}{\sigma_1}r_{12} + b_3\frac{\sigma_3}{\sigma_1}r_{13} + \ldots$ und entsprechend für die Korrelation

$$r_{z^*x_1} = \frac{\sigma_{z^*x_1}}{\sqrt{\sigma_1^2\sigma_{z^*}^2}} = b_1\frac{\sigma_1}{\sigma_{z^*}} + b_2\frac{\sigma_2}{\sigma_{z^*}}r_{12} + b_3\frac{\sigma_3}{\sigma_{z^*}}r_{13} + \ldots \quad . \tag{10}$$

Man beachte, dass es sich bei Gl. 9 im Unterschied zu Gl. 4 und 5 um ein deterministisches Modell (ohne eine Zufallsvariable u_1 oder u_2) handelt. Wie man sieht korreliert bei einer gegebenen Anzahl m von EVn eine EV, wie etwa x_1 um so mehr mit z*, der gewogenen Summe von $x_1, x_2, \ldots, x_m$, je stärker x_1 korreliert ist mit den anderen Variablen x_2, x_3 usw. (stets positive Korrelationen vorausgesetzt).

---

[35] Daran, dass die PSM als Verfahren selbst höchst unbefriedigend ist, ändern diese Überlegungen jedoch nichts.

Wie sieht es bei Hinzunahme von EVn, also etwa beim Übergang von m = 2 zu m = 3 aus? Angenommen es seien alle EVn standardisiert d.h. Mittelwert 0, Varianz 1 und es werde ein ungewogenes Mittel z* gebildet dann ist bei m = 2 und $b_1 = b_2 = 1/2$ sowie $\sigma_1^2 = \sigma_2^2 = 1$

(8a) $$\sigma_{z*}^2 = \frac{1}{4}(\sigma_1^2 + \sigma_2^2 + 2\sigma_{12}) = (1 + r_{12})/2 \text{ und}$$

(9a) $$\sigma_{z*x_1} = b_1\sigma_1^2 + b_2\sigma_{12} = (\sigma_1^2 + r_{12}\sigma_1\sigma_2)/2 = (1 + r_{12})/2, \text{ so dass gilt}$$

(10a) $$r_{z*x_1} = \sqrt{(1 + r_{12})/2} = r_{z*x_2}$$

Ganz analog erhält man bei m = 3 und $b_1 = b_2 = b_3 = 1/3$

(8b) $$\sigma_{z*}^2 = \left(\frac{1}{3}\right)^2 (\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2r_{12}\sigma_1\sigma_2 + 2r_{13}\sigma_1\sigma_3 + 2r_{23}\sigma_2\sigma_3) = \frac{3 + 2(r_{12} + r_{13} + r_{23})}{9} \text{ und}$$

(9b) $$\sigma_{z*x_1} = (b_1\sigma_1^2 + b_2\sigma_{12} + b_3\sigma_{13})/3 = (1 + r_{12} + r_{13})/3 \text{ und somit}$$

(10b) $$r_{z*x_1} = \frac{1 + r_{12} + r_{13}}{\sqrt{3 + 2(r_{12} + r_{13} + r_{23})}} \text{ und entsprechend } r_{z*x_2} = \frac{1 + r_{12} + r_{23}}{\sqrt{3 + 2(r_{12} + r_{13} + r_{23})}}.$$

Dieser Ausdruck wird i.d.R. geringer sein als $r_{z*,1}$ gem. Gl. 10a, insbesondere dann, wenn die bisherigen Variablen x_1 und x_2 mit der hinzugekommenen Variable x_3 nicht korrelieren, also $r_{13} = r_{23} = 0$. Im Fall unkorrelierter EVn erhält man gem. Gl. 10a bei zwei EVn $r_{z*x_1} = r_{z*x_2} = \sqrt{1/2} = 0{,}707$ und bei drei $r_{z*x_1} = r_{z*x_2} = r_{z*x_3} = 1/\sqrt{3} = 0{,}577$. Nimmt man dagegen eine hohe positive Korrelation an, so gilt bei einem z* als ungewogenes arithmetisches Mittel bei m = 2 $r_{z*x_i} = \sqrt{0{,}9} = 0{,}949$ (i = 1,2) und bei m = 3 $r_{z*x_i} = 2{,}4/\sqrt{7{,}8} = 0{,}931$ (mit i = 1,2,3).

b) Generalfaktormodell (Faktorenanalyse mit einem Faktor)

Man könnte mit der Faktorenanalyse prüfen, ob den m Indikatoren (EVn oder manifesten Variablen) überhaupt *ein* (latenter) "Faktor" zugrunde liegt, wie implizit bei der PSM unterstellt wird, oder ob es mehrere Faktoren sind[36]. Im Falle *eines* Faktors y müßte für je zwei Indikatoren x_i und x_j gelten

(11) $$r_{x_i x_j} = r_{yx_i} r_{yx_j} \text{ oder einfach } r_{ij} = r_{yi} r_{yj}.$$

Daraus folgt, dass z.B. für bei (je) drei EVn x_1, x_2, x_3 gelten muß[37]

(12) $$r_{1y} = \sqrt{\frac{r_{12}r_{13}}{r_{23}}},$$

was im Falle einer Punktsumme z anstelle von y zu vergleichen wäre mit

(10c) $$r_{z*1} = \frac{1}{\sigma_{z*}}(b_1\sigma_1 + b_2\sigma_2 r_{12} + b_3\sigma_3 r_{13}),$$

und was dann auch bei Berücksichtigung einer vierten Variablen x_4 sich nicht zu sehr unterscheiden sollte von $\sqrt{\frac{r_{12}r_{14}}{r_{24}}}$ und $\sqrt{\frac{r_{13}r_{14}}{r_{34}}}$. Wie man sieht liefert die Anwendung eines sol-

[36] Für den o. g. Lebensstandard Atlas vgl. v. d. Lippe, a.a.O. (und die dort angegebene Arbeit von M. Zwick).
[37] Die Korrelation r_{1y} ist die "Ladung" von x_1 im (einen) "Generalfaktor".

chen Modells, wie das des "Generalfaktors" ein empirisch überprüfbares Kriterium dafür, ob die Annahme, dass die Indikatoren eine und nur eine latente Dimension messen, überhaupt angemessen ist. Multipliziert man bei m = die Korrelationen r_{z*1} und r_{z*2} gem. Gl. 10 so erhält man mit

$$(13)\qquad r_{z*1}r_{z*2} = \frac{(b_1\sigma_1 + b_2\sigma_2 r_{12})(b_1\sigma_1 r_{12} + b_2\sigma_2)}{\sqrt{b_1^2\sigma_1^2 + 2b_1b_2\sigma_{12} + b_2^2\sigma_2^2}} = r_{12} + \frac{(1-r_{12})^2 B}{A+2B}$$

mit $A = g_1^2\sigma_1^2 + g_2^2\sigma_2^2$ und $B = g_1g_2\sigma_{12}$, oder

$$(14)\qquad r_{z*1}r_{z*2} = r_{12} + \frac{(1-r_{12})^2 g_1g_2\sigma_1\sigma_2}{(g_1\sigma_1 + g_2\sigma_2)^2},$$

was in der Regel nicht gleich r_{12} ist, wie dies im Falle von Gl. 11 gilt, also mit dem "Faktor" y statt der Summenvariable z*.

3. Reproduzieren von Strukturen (Punktsumme und Cluster)

Wir wollen im Folgenden mit zwei einfachen Zahlenbeispielen bei jeweils m = 2 EVn zeigen, dass es wenig Sinn machen kann, eine einfache Punktsumme zu bilden, wenn die Daten bereits in Form von Clustern vorliegen. Die Summenbildung kann dann nämlich dazu führen, dass praktisch alle Einheiten auf (annähernd) die gleiche Punktsumme kommen und so (mit der Punktsumme) die Unterscheidung in vorgegebene Cluster nicht mehr sichtbar wird.

Beispiel 1

Gegeben seien sechs Einheiten, wobei die ersten drei zur Gruppe (zum Cluster) 1 die Einheiten 4 bis 6 zur Gruppe 2 gehören. An den Einheiten wurden jeweils zwei Merkmale x_1 und x_2 beobachtet:

Einheit	x_1	x_2	$x_1 + x_2$
1	4	16	20
2	5	14	19
3	6	12	18
4	12	6	18
5	14	5	19
6	16	4	20

Es ist unschwer zu sehen, warum man hier von zwei Clustern sprechen kann (die Einheiten 1 bis 3 befinden sich links oberhalb der Winkelhalbierenden und die Einheiten 4 bis 6 rechts unterhalb (vgl. Abb. 2). Die Punktsummen nehmen die Werte 18, 19 und 20 an, sind also sehr ähnlich.

Die Summen verhalten sich wie die Projektionen auf die Winkelhalbierende, die mit dem Vektorprodukt $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \cos 45° \\ \cos 45° \end{bmatrix} = \frac{1}{2}\sqrt{2}$ (x_1 + x_2) gegeben sind, also mit der 0,707 fachen Punktzahl (die Strecke von Punkt [Einheit] 1 und von Punkt 6 zur Winkelhalbierende ist 0,707·20 = 14.14 usw.).

Äquivalent wären danach die Einheiten 1 und 6, 2 und 5 sowie 3 und 4. Ganz anders sieht es dagegen aus, wenn man die euklidischen Distanzen betrachtet. Die Distanzmatrix enthält die Werte der auf der nächsten Seite folgenden Tabelle.

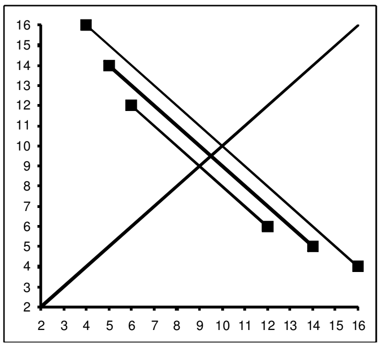
Zur Erklärung der Berechnung greifen wir zwei Werte heraus, nämlich

$$d_{12} = \sqrt{5} = \sqrt{(4-5)^2 + (16-14)^2} = 2{,}24 \text{ und}$$

$$d_{15} = \sqrt{221} = \sqrt{(4-14)^2 + (16-5)^2} = \sqrt{100+121} = 14{,}87$$

Die Distanzmatrix zeigt ebenfalls, dass es offenbar - wie auch dem Beispiel zugrunde gelegt - zwei Gruppen gibt, weil die Abstände *innerhalb* einer Gruppe zwischen 2,24 und 4,47 schwanken, *zwischen* den Gruppen aber erheblich größer sind (mit Werten zwischen 8,48 und 16,97).

Abb. 2



	2	3	4	5	6
1	2,24	4,47	12,81	14,87	16,97
2		2,24	10,63	12,73	14,87
3			8,48	10,63	12,81
4				2,24	4,47
5					2,24

Mit der metrischen multidimensionalen Skalierung könnte man prüfen, ob es eine und nur eine Skala (Dimension) gibt, auf der sich die sechs Objekte so aneinander reihen lassen, dass ihre Abstände auch tatsächlich den Abständen der Distanzmatrix entsprechen. Man sieht, dass man auch bei dieser Betrachtung, wenn man also von der Möglichkeit einer Struktur in Gestalt von Clustern ausgeht, Methoden besitzt um vor einer einfachen Punktsummenbildung empirisch zu prüfen, ob die PSM überhaupt Sinn macht.

Man kann neben den Distanzen auch die Diskriminanzfunktion für dieses Beispiel betrachten, die sich ergibt mit $\tilde{x} = a_1x_1 + a_2x_2 = x_2 - x_1$, so dass $a_1 = -1$ und $a_2 = +1$. Die Trennlinie $\tilde{x} = 0$ erlaubt eine eindeutige Zuordnung. Positive Werte von $\tilde{x}$ bedeuten Zugehörigkeit zur Gruppe 1, und negative zur Gruppe 2. Im Unterschied zur Punktsumme ergibt die Diskriminanzfunktion (ebenfalls eine Linearkombination, jedoch nicht mit den Gewichten + 1 und + 1 sondern -1 und +1) also eine erhebliche Unterschiedlichkeit, insbesondere auch der Einheiten (Objekte) 1 und 6 deren Werte + 12 und -12 sind, die aber beide die gleiche Punktsumme 20 haben. Man erhält die folgenden Distanzen aufgrund der Werte der Diskriminanzfunktion (und in Klammern aufgrund der Punktsummen)

	2	3	4	5	6
1	3 (1)	6 (2)	18 (2)	21 (1)	24 (0)
2		3 (1)	15 (1)	18 (0)	21 (1)
3			12 (0)	15 (1)	18 (2)
4				3 (1)	6 (2)
5					3 (1)

Wie man sieht spiegeln die Distanzen aufgrund der Diskriminanzfunktion $\tilde{x} = x_2 - x_1$ ziemlich genau die gleichen Größenverhältnisse wider, wie sie aufgrund der Distanzen im Streuungsdiagramm herrschen, während sich "nah" und "fern" aufgrund der Punktsumme ganz anders verhalten.

Das Versagen der Punktsumme kann, so könnte man argumentieren, darauf beruhen, dass – wie auch das Streuungsdiagramm zeigt - x_1 und x_2 negativ miteinander korreliert sind (es ist in der Tat r_{12} = - 0,9848)

Beispiel 2

Hier zeigt sich bei einer noch etwas stärkeren negativen Korrelation (r_{12}= -0,9981) dass die einfache Punktsumme Unterscheidungen erneut verwischt. Sieben der acht Einheiten, von denen die ersten drei zur Gruppe1 und die restlichen fünf zur Gruppe 2 gehören, haben die gleiche Punktzahl, nämlich 14.

Die Diskriminanzfunktion $\tilde{x} = 3{,}2774x_1 + 5{,}7290x_2$ zeigt demgegenüber deutliche Unterschiede und spiegelt auch durchaus die Gruppenzugehörigkeit wider, wie sie dem konstruierten Beispiel zugrunde liegt. Auch hier versagt die PSM wegen der negativen Korrelation zwischen den beiden EVn.

Einheit	x_1	x_2	$x_1 + x_2$	Diskriminanzfkt.
1	2	12	14	75,30
2	1	13	14	77,75
3	3	11	14	72,85
4	12	2	14	50,79
5	13	1	14	48,34
6	11	3	14	53,24
7	10	4	14	55,69
8	14	1	15	51,61

Einheiten der ersten Gruppe haben $\tilde{x}$ - Werte zwischen 72,85 und 77,75 und sind damit deutlich verschieden von solchen der zweiten Gruppe, bei denen sich die $\tilde{x}$ – Werte um 50 herum bewegen. Nach der Punktsumme beurteilt müßten die Einheiten 2 und 5 genauso "gleich" sein wie etwa 4 und 5, während die Abstände auf der $\tilde{x}$ - Skala 29,41 und 2,45 betragen (und die euklidischen Distanzen im Streuungsdiagramm sind 16,97 und 1,41). Es ist auch nicht zu er *zwei*-dimensionalen Betrachtung kennen, warum die Einheit 8 (mit der Punktsumme 15 statt 14) sich von anderen erheblich unterscheidet. Die Diskriminanzfunktion zeigt das nicht und auch die Abstände entsprechen dem nicht (Einheit 8 ist z.B. näher an Einheit 5 als Einheit 5 an Einheit 3 usw.).

Es zeigt sich also dass das bereits genannte "Kompensieren" geringer Werte in einer Dimension durch große Werte in einer anderen Dimension, was bei der PSM allein wegen der Summenbildung einstellt, von Schaden sein kann, wenn Cluster vorliegen, also deutliche Gruppenunterschiede bei einer *zwei* Dimensionen, die bei einer künstlich *ein*-dimensionalen Betrachtung verschwinden.

4. Die PSM bei Hinzunahme oder Austausch eines Merkmals

(Dieser Abschnitt wurde von A. Kladroba verfasst)

Ein einfacher Zugang um die Nützlichkeit einer bestimmten Methodik zu analysieren, ist nach dem Verhalten der Ergebnisse bei bestimmten Datenänderungen zu fragen. Es handelt sich also um eine Art axiomatischer Zugang. Allerdings soll an dieser Stelle kein komplettes Axiomensystem erarbeitet, sondern nur einige Plausibilitätsüberlegungen vorgetragen werden.

Wir wollen im Folgenden die Frage stellen, wie sich der mit Hilfe der PSM ermittelte Index ändert, wenn

a) ein Merkmal hinzukommt
b) ein Merkmal gegen ein anderes ausgetauscht wird.

Wir werden dazu exemplarisch acht Szenarien betrachten, die sich neben den angeführten Punkten a) und b) aufgrund folgender Unterscheidungen bilden.

1. Art der Gewichte mit

a) „echten“ Gewichten, d.h. dass die Gewichte die Eigenschaften $0 \leq b_i \leq 1$ und $\sum b_i = 1$ haben
b) „unechten“ Gewichten, d.h. das die Gewichte diese Eigenschaften nicht haben. Der Einfachheit halber wollen wir in diesem Fall ganzzahlige Gewichte annehmen.[38]

2. Anzahl der Merkmalsträger (Länder, Unternehmen usw.)

a) ein Merkmalsträger, wobei hier die Fragestellung lauten wird, ob sich der Index bei der Hinzunahme bzw. dem Austausch eines Merkmals auf die vorher zu erwartende Art entwickelt
b) mehrere (hier: zwei) Merkmalsträger, wobei hier die Frage nach der Trennfähigkeit der Indizes im Vordergrund steht.

Aus diesen Unterscheidungen ergeben sich folgende Szenarien:

	ein Merkmalsträger		zwei Merkmalsträger	
	echte Gewichte	unechte Gewichte	echte Gewichte	unechte Gewichte
Hinzunahme eines Merkmals	Szenario 1	Szenario 2	Szenario 5	Szenario 6
Austausch eines Merkmals	Szenario 3	Szenario 4	Szenario 7	Szenario 8

Wir gehen bei den folgenden Betrachtungen davon aus, dass die ursprünglichen Messdaten bereits in eine Punkteskala (in den folgenden Beispielen mit $0 \leq x \leq 1000$, wobei null den schlechtesten und 1000 den besten Wert repräsentieren) transformiert worden sind.

Szenario 1

Die Besonderheit dieses Szenarios (z.B. im Vergleich zu den Szenarien 2 oder 3) besteht darin, dass aufgrund der Restriktion, dass die Summe der Gewichte immer den Wert 1 ergeben soll, beim Hinzufügen eines Merkmals eine Umgewichtung auch der schon vorhandenen Merkmale notwendig ist. In Anlehnung an die bekannten Mittelwertaxiome[39] (hier das Axiom der Ergänzung) ist folgendes Verhalten der Punktsumme zu erwarten:

- $z^*_{n+1} < z^*_n$, wenn $x_{n+1} < z^*_n$
- $z^*_{n+1} > z^*_n$, wenn $x_{n+1} > z^*_n$
- $z^*_{n+1} = z^*_n$, wenn $x_{n+1} = z^*_n$

Besonders der letzte Punkt (unveränderte Punktsumme, wenn der neu hinzugefügte Merkmalswert der alten Punktsumme entspricht) scheint uns für eine nähere Betrachtung sehr geeignet. Zur Illustration möge ein Beispiel dienen:

[38] Die ungewogene Punktsumme z ist hiervon ein Spezialfall.
[39] vgl. z.B. von der Lippe (1993), Deskriptive Statistik, Stuttgart, S. 43ff.

Wir gehen von ursprünglich zwei Merkmalen mit den Ausprägungen $x_1 = 100$ für das Merkmal 1 und $x_2 = 600$ für das Merkmal 2 aus. Diese werden mit $b_1 = 0{,}6$ und $b_2 = 0{,}4$ gewichtet, woraus sich eine gewogene Punktsumme von $z_2^* = 300$ ergibt. Wir fügen ein weiteres Merkmal mit der Ausprägung $x_3 = 300$ hinzu. Wir erwarten also, dass sich der Index nicht ändert. Fraglich ist jetzt, welche Gewichte zu wählen sind um diese Ziel zu erreichen. Es ist leicht einsehbar, dass bei der Wahl der „falschen“ Gewichte, der Index sowohl steigen als auch sinken kann. Z. B. fällt der Index bei $b_1 = 0{,}6$, $b_2 = 0{,}3$ und $b_3 = 0{,}1$ auf 270. Um den gleichen Indexwert wie vorher zu erhalten müssen die neuen Gewichte derart festgelegt werden, dass die Quotienten die gleichen sind wie vorher. In unserem Beispiel hat sich für die ursprünglichen Gewichte der Quotient $\frac{b_1}{b_2} = \frac{0{,}6}{0{,}4} = 1{,}5$ ergeben. Wählt man nun $b_3 = 0{,}1$ müssen die anderen Gewichte bei einer Summe von 0,9 die Werte $b_1 = 0{,}54$ und $b_2 = 0{,}36$ annehmen. Der Quotient ist somit $\frac{b_1}{b_2} = \frac{0{,}54}{0{,}36} = 1{,}5$ und der Index $z_3^* = 300$. Damit sind im übrigen auch die beiden anderen Verhaltensregeln für steigende bzw. fallende Indizes, wenn $x_{n+1} < z_n^*$ bzw. $x_{n+1} > z_n^*$ gilt, sichergestellt. Dies gilt natürlich auch für mehr als zwei bzw. drei Merkmale. Fügt man unserem Beispiel noch ein viertes Merkmal mit $x_4 = 300$ und $b_4 = 0{,}2$ hinzu, so dass wiederum ein konstanter Index zu erwarten ist, dann sind die restlichen Gewichte mit $b_1 = 0{,}432$, $b_2 = 0{,}288$ und $b_3 = 0{,}08$ zu wählen, damit der Index unverändert bleibt.

Szenario 2

Bei „unechten“, d.h. hier ganzzahligen Gewichten, die sich naturgemäß nicht zu eins addieren, ist bei der Hinzunahme eines weiteren Merkmals immer zu beobachten, dass der Index steigt, was nicht weiter überrascht. Allerdings sollte man sich die Frage stellen, ob dies ein wünschenswertes Verhalten ist, da ein Anstieg des Indexes auch dann zu beobachten ist, wenn das neue Merkmal eine äußerst „schlechte“, also niedrige Ausprägung, hat. Die Interpretationsfähigkeit der Punktsumme als absolute Zahl (also nicht im Vergleich zu den Punktsummen anderer Merkmalsträger) ist vor diesem Hintergrund doch sehr zweifelhaft.

Szenarien 3 und 4

In den Szenarien 3 und 4 werden die Gewichte wie unter 1. bzw. 2 gewählt und ein Merkmal gegen ein anderes ausgetauscht. Dabei wird unterstellt, dass die ursprüngliche Gewichtung beibehalten wird. Es ist sicherlich leicht einsichtig, dass in beiden Fällen der Index steigt, wenn der neue Wert größer ist als der alte und im umgekehrten Fall entsprechend sinkt.

Szenario 5

In den Szenarien 5 - 8 werden zwei Merkmalsträger betrachtet. Im Gegensatz zu den Szenarien 1 - 4 interessiert hier weniger die absolute Entwicklung der einzelnen Indizes, sondern vielmehr deren Differenz als Zeichen für die Trennfähigkeit der PSM. Wir werden dabei im folgendem davon ausgehen, dass bei beiden Merkmalsträgern die gleichen Gewichte angenommen werden.

Im folgenden Beispiel werden wir die Situationen unterscheiden, dass das neu hinzugenommene Merkmal offensichtlich

a) nicht zur Unterscheidung der Merkmalsträger beiträgt
b) deutlich zur Unterscheidung beiträgt.

Beispiel a)

Die Ausgangssituation sieht wie folgt aus:

	x_1	x_2	PS
MT 1	100	200	140
MT 2	900	800	860
Gewichte	0,6	0,4	

MT = Merkmalsträger; PS = Punktsumme

Das hinzukommende dritte Merkmal sei bei beiden Merkmalsträgern identisch und trägt somit nicht zur Unterscheidung der Objekte bei. Wie ist also jetzt die Gewichtung b_1 und b_2 zu wählen, wenn $b_3 = 0{,}1$ festgesetzt wird?

Zunächst ist festzustellen, dass absolute Verhalten der beiden Punktsummen maßgeblich vom Wert des dritten Merkmals abhängt:

- Ist $x_3 < 500$ kann z_1^* zu- oder abnehmen, z_2^* nimmt auf jeden Fall ab.
- Ist $x_3 > 500$ gilt das Umgekehrte.

Die entscheidende Frage sollte an dieser Stelle ja aber sein: Wie verhält sich die Differenz zwischen den Indizes? Hier ist festzustellen: Die Differenz nimmt auf jeden Fall ab. Die Stärke der Abnahme ist abhängig von den Gewichten. Eine Zunahme ist in diesem Beispiel nicht möglich. Ein Extremfall liegt dann vor, wenn $b_1 = 0{,}9$ ist und somit das zweite Merkmal also gar nicht beachtet wird. Dann ist die Differenz wie bei der Betrachtung der Merkmale 1 und 2.

Als Ergebnis kann festgehalten werden: Das dritte Merkmal sorgt auf jeden Fall für eine Verschlechterung der Unterscheidung, was bei den Vorüberlegungen ja auch erwartet worden ist.

Beispiel b):

Als Ausgangspunkt diene der von Beispiel a). Es wird ein Merkmal hinzugefügt, das die Unterscheidung der Merkmalsträger auf jeden Fall erhöhen sollte mit $x_{31} = 100$ und $x_{32} = 900$ und wiederum mit $b_3 = 0{,}1$. Dann lässt sich zeigen, dass für $b_1 > 0{,}5$ (und somit $b_2 < 0{,}4$) die Differenz zwischen den Indizes zu- und im umgekehrten Fall abnimmt. Die Trennung der beiden Merkmalsträger kann also je nach Gewichtung besser oder schlechter werden. Anders als in Szenario 1 spielt dabei der Quotient der Gewichte offensichtlich keine Rolle.

Szenario 6

Verwendet man für die Betrachtung des Szenarios 5 „unechte" Gewichte, sind folgende Ergebnisse leicht nachzuvollziehen:[40]

1. Nimmt das dritte Merkmal für beide Merkmalsträger den gleichen Wert an, steigt die Punktsumme für beide Merkmalsträger um den gleichen Betrag (nämlich um ein ganzzahliges Vielfaches des neuen Merkmalswertes), so dass die Differenz zwischen den Indizes unverändert bleibt. D.h. dass, obwohl das neue Merkmal nicht differenziert, dies durch die absoluten Werte der Indizes nicht angezeigt wird. Bildet man dagegen die Quotienten $Q_1^{alt} = \frac{\left|z_1^{alt} - z_2^{alt}\right|}{z_1^{alt}}$ und $Q_1^{neu} = \frac{\left|z_1^{neu} - z_2^{neu}\right|}{z_1^{neu}}$ mit z_1^{alt} dem ursprüng-

[40] Anmerkung: Nach Hinzunahme des dritten Merkmals bleiben die Gewichte für die ersten beiden Merkmale unverändert.

lichen Index für Merkmalsträger 1 bzw. z_2^{alt} dem ursprünglichen Index für Merkmalsträger 2 und den entsprechenden neuen Punktsummen z_1^{neu} und z_2^{neu}, dann gilt $Q_1^{neu} < Q_1^{alt}$.[41] Es liegt also eine Abnahme der Unterscheidung zwischen den Merkmalsträgern vor.

2. Differenziert das neue Merkmal dagegen gut und zwar im Sinne der Merkmale 1 und 2, steigt auch die Differenz in Abhängigkeit vom Gewicht.

Szenarien 7 und 8

Der Austausch eines Merkmals gegen ein anderes bewirkt unabhängig, ob „echte“ oder „unechte“ Gewichte verwendet werden, dass die Unterscheidung der Ratingobjekte besser wird, wenn das neue Merkmal besser als das alte differenziert und umgekehrt. Dabei werden - wie bereits erwähnt - die ursprünglichen Gewichte beibehalten.

Zusammenfassung des Kapitels

Aus unseren Überlegungen lässt sich zunächst das etwas simpel scheinende Fazit ziehen, dass das „richtige“ im Sinne eines sinnvollen Verhaltens einer Punktsumme fast ausschließlich von der Wahl der Gewichte abhängt. Dabei lässt sich unterscheiden:

1. Der Austausch eines Merkmals gegen ein anderes (Szenarien 3 und 4 sowie 7 und 8) ist unproblematisch solange die ursprünglichen Gewichte beibehalten werden.
2. Ein eindeutiges Verhalten zeigt der Index auch bei ganzzahligen Gewichten (Szenario 2). Allerdings ist hier zweifelhaft, ob dieses Verhalten wirklich wünschenswert ist.
3. Mit Hilfe einer einfachen Regel (Quotientenkonstanz) lässt sich die Erfüllung des Ergänzungsaxioms für Mittelwerte erreichen (Szenario 1).
4. Problematisch ist offensichtlich die Beurteilung der Trennfähigkeit der PSM. Während in Szenario 6 durch die Wahl einer relativen Differenz als Maßstab noch eine eindeutige Aussage zu erreichen ist, erscheint dies in Szenario 5 nicht mehr möglich. Man sollte dabei betonen, dass die Hinzunahme oder der Wegfall eines Merkmals sicherlich keine ausschließlich theoretische Überlegung sondern im Gegenteil von hoher Praxisrelevanz ist, so dass hier neben den bereits skizzierten Schwächen ein weiterer schwerer Nachteil der PSM zu erkennen ist.

5. Fazit

Die vorausgegangenen Ausführungen haben gezeigt, dass wir die Punktsummenmethode für eine ungeeignete Form der Messung komplexer Variablen halten. Dem einzigen (vermeintlichen) Vorteil, nämlich ihrer Einfachheit in der Anwendung und dem intuitiven Verständnis, das man ihr entgegenbringt (das aber gleichzeitig auch zu unangebrachten Interpretationen verführt) stehen die folgenden eklatanten Nachteile gegenüber:

1. Das Fehlen jeglicher Theorie, die das Verfahren erklären und unterstützen würde,
2. die Gefahr offensichtlich unplausibler Ergebnisse und
3. ein hoher Anteil recht willkürlicher Annahmen bezüglich der Wahl der Gewichte, die zu höchst unterschiedlichen Ergebnissen führen.

Die Statistik hält hier sicherlich im Rahmen der multivariaten Analyse Verfahren bereit, die zwar weniger intuitiv und methodisch anspruchsvoller sind, deren Ergebnisse aber letztlich auf einem deutlich stärkeren Fundament ruhen und damit für den Anwender von viel größerem Nutzen sind.

---

[41] Man bekommt natürlich das gleiche Ergebnis, wenn im Nenner die Indizes des zweiten Merkmalsträgers stehen.