

Sören Werner, Kristina Kraus

Entwicklung eines Sozialindex mittels einer explorativen Hauptkomponentenanalyse

Der Freiburger Sozialindex wurde anhand einer explorativen Hauptkomponentenanalyse entwickelt und dient dazu, die Multidimensionalität der sozialen Lage in den (statistischen) Bezirken auf einen Wert herunterzubrechen. Das zur Ermittlung verwendete multivariate Analyseverfahren erlaubt es, valide Vergleichswerte von Bezirken untereinander und Entwicklungen im Zeitverlauf abzubilden, und ist stabil genug, um nicht durch Ausreißer oder statistische Fehler beeinflusst zu werden. Der Sozialindex wird mittels sieben Indikatoren gemessen, die durch ein iteratives Verfahren aus insgesamt 130 Indikatoren ausgewählt worden sind.

Einleitung

Es gab zwei konkrete Anlässe, den Sozialindex ausgerechnet im Jahr 2022 zu entwickeln. Der erste war der Wunsch des Amtes für Soziales, soziale Benachteiligung zu messen und die Datengrundlage für die Quartiersarbeit zu aktualisieren. Dadurch wird ermittelt, welche Stadtgebiete hier stärker in den Fokus rücken können. Der zweite Anlass war eine bedeutende datentechnische Voraussetzung: die breite Verfügbarkeit von geeigneten Indikatoren aus verschiedenen Themengebieten und Datenquellen. Nach vielen Jahren des Aufschiebens wurde 2021 endlich das Indikatorenset der Freiburger Statistikstelle neu aufgesetzt: eine halbautomatisierte Zusammenstellung von Kennzahlen aus allen verfügbaren Datenquellen auf drei Raumebenen (Gesamtstadt, Stadtbezirke, statistische Bezirke) mit langen Zeitbezügen (teilweise zurückgehend bis in das Jahr 1990). Diese Datenstruktur ermöglicht es, mit wenig Aufwand vielfältige statistische Analysen anzustoßen.

Die Idee einer explorativen Hauptkomponentenanalyse

Es war eine bewusste Entscheidung, die Ermittlung eines Sozialindex primär anhand statistischer Kriterien durchzuführen und sich nicht von theoretischen Vorüberlegungen oder Vorgaben beeinflussen zu lassen. Die Idee, die Analyse unvoreingenommen mit allen verfügbaren Indikatoren zu beginnen und sich nach und nach dem Ergebnis zu nähern, konnte am besten mit einer explorativen Hauptkomponentenanalyse umgesetzt werden. Demzufolge wurden weder andere Fachämter noch andere externe, nicht-statistische Kriterien hinzugezogen. Die Auswahl der konstituierenden Indikatoren sollte durch das statistische Modell erfolgen. Selbstverständlich schließt dies nicht aus, dass eine fundierte Kenntnis über die in das Modell einfließenden Indikatoren, darüber, was diese messen und wie die Daten zu interpretieren sind, nicht nur vorhanden sondern auch als Voraussetzung gelten kann.

Die technische Idee:

Das Ziel der explorativen Hauptkomponentenanalyse ist es, auf Grundlage der vorliegenden Daten die Beziehungen zwischen den Variablen mit einer möglichst geringen Anzahl an Dimensionen zu reproduzieren oder um auf latente Variablen („Faktoren“) zu schließen, ohne dabei Vorgaben zu den Bezie-


Sören M. Werner

(M.A. Soziologie) Sachgebietsleiter Kommunalstatistik, Stadt Freiburg i. Br.

 soeren.werner@stadt.freiburg.de

Kristina Kraus

(M.A. Soziologie, Politikwissenschaften) Mitarbeiterin Kommunalstatistik, Stadt Freiburg i. Br.

 kristina.kraus@stadt.freiburg.de

Schlüsselwörter:

Index – Indikatoren – Sozialindex – Faktorenanalyse – Hauptkomponentenanalyse – multivariate Statistik – Soziale Benachteiligung

hungen der Variablen untereinander zu machen (wie bei der konfirmatorischen Faktorenanalyse). Die Anzahl der zu extrahierenden Faktoren ist unbekannt (Backhaus et al. 2021¹: 417). Es wird versucht, mit wenigen Komponenten die maximale Information/Varianz der Variablen (maximal = Anzahl der in Analyse einbezogenen Variablen) zu reproduzieren (Backhaus et al. 2021: 435, 440).

Die inhaltliche Idee:

Eine einzelne Zahl, ein Index, soll möglichst adäquat die soziale Wirklichkeit und die Multidimensionalität der sozialen Lage wiedergeben bzw. die relativen Unterschiede zwischen den verschiedenen städtischen Gebieten deutlich machen. Dafür sollen in allen Daten geschaut werden, welche sich als geeignet darstellen, eine „Gemeinsamkeit“ abzubilden.

Dieser Ansatz erlaubt es, durch ein iteratives Verfahren nach und nach Indikatoren in einer Weise zu „eliminieren“, dass sich eine Komponente herauskristallisieren lässt und dabei die Modellgüte stetig verbessert wird.

Am Ende soll ein Set von Indikatoren stehen, die alle eine hohe Faktorladung aufweisen, durch verschiedene Modellmaße als geeignet befunden wurden und dabei trotzdem unterschiedliche Dimensionen der sozialen Lage abbilden.

Die Voraussetzung für eine explorative Hauptkomponentenanalyse

Infolge dessen wurden die vorliegenden 130 Indikatoren der Bereiche Arbeitsmarkt, Bevölkerungsbestand und -bewegungen, Fläche, Gebäudebestand, Haushalte, Kraftfahrzeuge, Soziales, Wahlen und Wohnungsmarkt analysiert. Es handelt sich dabei um klassische Kennzahlen wie z. B. Arbeitslosenquote, Beschäftigtenquote, Haushaltsquoten, Ausländeranteil, verschiedene Altersquotienten, diverse Geburtenziffern, Quoten von staatlichen Transferleistungsempfänger*innen sowie Wahlbeteiligung und Wohndichte, aber auch speziellere Kennzahlen wie Neuwagenanteile (Pkw), Religionszugehörigkeiten, Briefwahlanteile, Anteile selbstgenutzten Wohneigentums, Räume je Wohnung, Scheidungsquote und Stimmenanteile von Parteien.

Im Grunde ist der gesamte kleinräumig verfügbare Datenbestand der Statistikstelle abgebildet. Das Set soll laufend überprüft und erweitert werden.

Zu Beginn der Ermittlung des Index mussten drei Entscheidungen getroffen werden:

- Raumbezug der Daten: hier war schnell klar, dass die Ebene der 167 statistischen Bezirke die geeignetste Ebene ist – groß genug für ausreichend Fälle, klein genug für eine ausreichende räumliche Differenzierung des Stadtgebiets.
- Zeitbezug der Daten: hier sollte ein möglichst aktueller Zeitbezug gewählt werden, um zum einen valide Werte für die Gegenwart zu haben und da zum anderen für vergangene Jahre nicht alle Indikatoren zur Verfügung stehen. Gleichzeitig sollten es möglichst „durchschnittliche“ Jahre ohne besondere Ereignisse sein. Daher wurde eine Zeit nach der Zuwanderung von Schutzsuchenden des Jahres 2015 und vor der Pandemie gewählt. Um Ausreißer einzufangen

und um strukturelle Zusammenhänge besser erkennen zu können, wurde ein 4-Jahres-Mittel (2016–2019) gewählt.

- Sachbezug der Daten bzw. Auswahl der Indikatoren: hier war die klare Entscheidung, zunächst alle Indikatoren in das Modell zu werfen, um ergebnisoffen zu denken und das Modell aufgrund statistischer Kennziffern entscheiden zu lassen, welche Indikatoren nach und nach ausgeschlossen werden.

Alle Indikatoren wurden mithilfe der z-Transformation zudem auf eine einheitliche Skala standardisiert, um nicht „Äpfel mit Birnen“ zu vergleichen. Das bedeutet beispielsweise, dass ein Wohnflächenwert von 40 Quadratmetern pro Person vergleichbar gemacht wird mit einer Arbeitslosenquote von 6,5 Prozent. Alle Indikatoren erhalten durch die durchgeführte Standardisierung einen Mittelwert von 0 und eine Standardabweichung von 1.

Das Ergebnis: ein stabiler, multidimensionaler Sozialindex

Unter Anwendung der Hauptkomponentenmethode verbleiben folgende sieben Indikatoren, die untereinander stark korrelieren (Abbildung 1) und hinter denen ein Faktor erkennbar ist:

- Arbeitslosenquote Frauen: Dies ist nicht nur ein Indikator für die allgemeine Erwerbsbeteiligung, sondern tangiert auch Status und Rolle der Frau.
- Durchschnittsalter der Mütter bei der Geburt: Hier steht die These im Raum, dass dies nicht nur ein Hinweis auf den Erwerbsstatus ist, sondern auch ein indirekter Bildungsindikator.
- Anteil Personen mit persönlichem Migrationshintergrund: Hatte höhere Korrelationen als der Ausländeranteil und deutet auf die kulturellen und ökonomischen Ressourcen hin, auch nach Erwerb der deutschen Staatsangehörigkeit.
- Quote Leistungsempfänger*innen nach SGB XII (Grundsicherung im Alter und bei Erwerbsminderung): klassischer Indikator von Altersarmut und der Erwerbsbiographie.
- Quote Personen in Bedarfsgemeinschaften nach SGB II (0 bis unter 65 Jahre): klassischer, indirekter Indikator des Erwerbs- und Einkommensstatus.
- Wahlbeteiligung bei Bundestagswahlen: bewährter Indikator für die politisch-gesellschaftliche Partizipation der deutschen (!) Bevölkerung.
- Wahlbeteiligung bei Landtagswahlen: komplementärer Indikator zur Bundestagswahl, auch um fehlende Jahre ohne Bundestagswahl abzudecken.

Die erklärte Gesamtvarianz des Sozialindex beträgt 77,2 Prozent und in Abbildung 2 sind die Komponentenladungen und Kommunalitäten der einzelnen Indikatoren zu sehen.

Zuletzt werden die Faktorwerte des Sozialindex für die statistischen Bezirke mithilfe einer Regressionsanalyse bestimmt. Da die Ausgangsindikatoren standardisiert in die Berechnung eingehen, sind auch die Faktorwerte standardisiert. Die Faktorwerte der statistischen Bezirke sind deren (fiktive) Lagewerte.

Abbildung 1: Korrelationsmatrix der ausgewählten Indikatoren für den Sozialindex

	Arbeitslosigkeit Frauen	Alter Mütter b. Geburt	Migrationshintergrund	SGB XII Grundsicherung	SGBII Personen in BG	Wahlbeteiligung BTW	Wahlbeteiligung LTW
Arbeitslosenquote Frauen		-0,62	0,72	0,71	0,94	-0,82	-0,72
Durchschnittsalter der Mütter bei Geburt	-0,62		-0,74	-0,49	-0,67	0,70	0,67
Anteil Personen mit persönlichem Migrationshintergrund	0,72	-0,74		0,63	0,78	-0,80	-0,76
Quote Leistungsempfänger*innen SGB XII (Grundsicherung)	0,71	-0,49	0,63		0,72	-0,73	-0,65
Quote Personen in Bedarfsgemeinschaften nach SGB II	0,94	-0,67	0,78	0,72		-0,83	-0,74
Wahlbeteiligung Bundestagswahlen	-0,82	0,70	-0,80	-0,73	-0,83		0,90
Wahlbeteiligung Landtagswahlen	-0,72	0,67	-0,76	-0,65	-0,74	0,90	

Es werden stets 3-Jahres-Mittel zur Analyse herangezogen, d.h. der Sozialindexwert des Jahres 2016 beinhaltet die Jahre 2014, 2015 und 2016. Somit wurde über alle statistischen Bezirke eine Zeitreihe von 2016 bis 2021 gebildet. Anhand derer können Veränderungen der sozialen Lage gemessen und dargestellt werden.

Wozu dient der Sozialindex?

Nun stellt sich die Frage: Wozu das Ganze? Tatsächlich war neben der konkreten Anfrage des Sozialamtes, einen Index zur räumlichen Einteilung und Erweiterung der Quartiersarbeit zu haben², eine große Portion stadtforscherischer Neugier im Spiel. Was kommt heraus, wenn wir alle verfügbaren Daten zusammen betrachten, welche Gemeinsamkeiten lassen sich finden, wo gibt es scheinbar wirkmächtige Faktoren, die hinter den vordergründigen Phänomenen stehen. Die Ausarbeitung der drei weiteren Indizes (siehe unten) war Ausdruck dieser Neugier und dieses Forschungsinteresses.

Unabhängig von der Freude an der Statistik stehen die Autor*innen trotzdem fachlich und inhaltlich voll hinter diesem Index: Er ist ein multidimensionales Maß der sozialen Lage und indiziert Gebiete mit Personen, die eine vergleichsweise geringe gesellschaftliche Teilhabe, eine „soziale Benachteiligung“, aufweisen, beobachtet deren Entwicklung und bietet Steuerungsbedarfe.

Deshalb ist es wünschenswert, dass möglichst viele Fachämter sich des Sozialindex annehmen und ihn nutzen. Aufgrund seiner Multidimensionalität ist er auch für verschiedene Ämter interessant: Amt für Migration und Integration, Amt für Soziales, Amt für Kinder und Jugend, Amt für Schule und Bildung, Stadtplanungsamt etc. Es ist an dieser Stelle jedoch schlicht zu früh, um hier eine Bewertung zur Akzeptanz und zum Einsatz in Planung und Steuerung machen zu können, aber es gibt bereits weitere Interessensanfragen.

Abbildung 2: Komponentenladung, Kommunalitäten und Anteil nicht erklärter Varianz der ausgewählten Indikatoren für den Sozialindex

	Komponentenladung	Kommunalitäten	Anteil Varianz, der nicht durch Faktor erklärt werden kann
Arbeitslosenquote Frauen	0,90	0,82	0,18
Durchschnittsalter der Mütter bei Geburt	-0,79	0,63	0,37
Anteil Personen mit persönlichem Migrationshintergrund	0,88	0,78	0,22
Quote Leistungsempfänger*innen SGB XII (Grundsicherung)	0,80	0,64	0,36
Quote Personen in Bedarfsgemeinschaften nach SGB II	0,93	0,86	0,14
Wahlbeteiligung Bundestagswahlen	-0,94	0,89	0,11
Wahlbeteiligung Landtagswahlen	-0,89	0,79	0,21

Abbildung 3: Sozialindexwerte der Freiburger statistischen Bezirke 2021

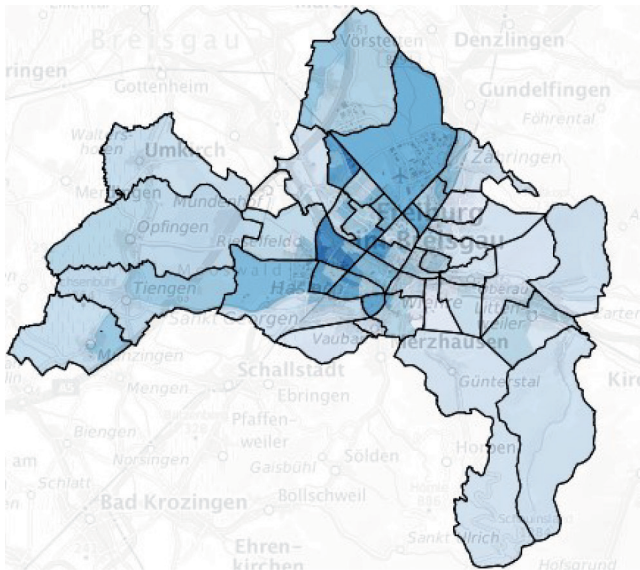


Abbildung 4: Komponentenladung, Kommunalitäten und Anteil nicht erklärter Varianz der ausgewählten Indikatoren für den Hochbetagtenindex

	Komponentenladung	Kommunalitäten	Anteil Varianz, der nicht durch Faktor erklärt werden kann
Greying-Index	0,91	0,82	0,18
Proportion der Alten zu den Jungen	0,78	0,60	0,40
Natürlicher Saldo	-0,91	0,83	0,17
Anteil Einpersonenhaushalte 75 Jahre und älter	0,89	0,79	0,21
Quote Leistungsempfänger*innen SGB XII (Hilfe zur Pflege)	0,74	0,55	0,45

Weitere Anwendungsmöglichkeiten derselben Methode: Hochbetagtenindex, Wohnungsneubauindex, Familienwohnindex

Im Zuge der explorativen Hauptkomponentenanalyse für den Sozialindex sind Komponenten aufgefallen, die auf weitere Faktoren hindeuteten. In derselben Methodik wurden somit weitere Indizes herausgearbeitet. Die Namensgebung spiegelt dabei den Versuch wider, deren inhaltliche Aussage ein wenig zu überschreiben.

Hochbetagtenindex

Der Hochbetagtenindex mit einer erklärten Gesamtvarianz von 72,0 Prozent besteht aus folgenden Indikatoren:

- Greying-Index – ein Maß, welches das Verhältnis von Personen im Alter von 80 Jahren und älter zu Personen zwischen 60 und 80 Jahren anzeigt.
- Proportion der Alten zu den Jungen – ebenfalls ein Altersmaß, welches das Verhältnis der 65-Jährigen und Älteren zu den unter 15-Jährigen anzeigt.
- Natürlicher Saldo – zeigt entweder einen Geburtenüberschuss oder einen Sterbefallüberschuss an.
- Anteil Einpersonenhaushalte 75 Jahre und älter – ein Maß zur Anzeige des Anteils von allein lebenden Senior*innen an allen Haushalten.
- Quote Leistungsempfänger*innen nach SGB XII (Hilfe zur Pflege) – die Quote zeigt an, wie hoch der Anteil in der Bevölkerung ist, der Hilfe zur Pflege nach SGB XII bezieht.

Der Hochbetagtenindex zeigt Gebiete mit vielen älteren Menschen an und wird erwartbar stark von der Präsenz von Seniorenwohnanlagen und Pflegeheimen beeinflusst. Diese Bezirke können jedoch im interaktiven Indikatorenatlas mittels eines Filters ausgeblendet werden, wodurch man eine etwas abgeschwächte Verteilung erhält.

Wohnungsneubauindex

Der Wohnungsneubauindex ist ein sehr volatiler Index mit starken jährlichen Schwankungen. Er zeigt Gebiete an, in denen neu gebaut wird oder wurde und sich dementsprechend auch die Personen- und Haushaltszahl verändert hat:

- Bevölkerungsentwicklung (Veränderungsrate) – die Veränderungsrate beschreibt die relative Bevölkerungszu- oder -abnahme zum Vorjahr.
- Anteil Neubauwohnungen letzte fünf Jahre – der Anteil von neu gebauten Wohnungen, die in den letzten fünf Jahren entstanden sind, am gesamten Wohnungsbestand.
- Veränderungsrate Wohnungsbestand Fertigstellungen – Veränderung des Wohnungsbestandes durch fertiggestellte Wohnungen zum jeweiligen Vorjahr.
- Veränderungsrate Haushalte – die Veränderungsrate beschreibt die relative Haushaltszu- oder -abnahme zum Vorjahr.

Der Wohnungsneubauindex könnte behilflich sein, Veränderungen in der Bevölkerungs- und Sozialstruktur, die sich in anderen Indikatoren und Indizes im Laufe der Zeit bemerkbar machen, im Nachhinein zu beziffern. Er weist eine erklärte Gesamtvarianz von 83,1 Prozent auf.

Familienwohnindex

Der Fokus beim Familienwohnindex liegt genauso stark auf dem Aspekt Wohnen wie auf dem Aspekt Familie. Es ist kein reiner Familienindex, sondern er beinhaltet weitere Indikatoren, die auf die Wohnumgebung hindeuten:

- Jugendquotient – ein klassisches Altersmaß, welches den Anteil der unter 15-Jährigen an der Gesamtbevölkerung anzeigt.
- Durchschnittliche Anzahl Räume je Wohnung – hiermit werden große Wohnungen bzw. Wohnungen mit vielen Räumen angezeigt.
- Durchschnittliche Haushaltsgröße – die Anzahl der Personen pro Haushalt korreliert sehr stark mit dem Familienanteil.
- Anteil Familienhaushalte – hier sind alle Haushalte mit Kindern enthalten.
- Pkw pro Haushalt – zeigt die Anzahl der Personenkraftwagen pro Haushalt und damit typische „Mittelschicht“-Familienwohngebiete an.

Der Familienwohnindex ist durch die verschiedenen Indikatoren stabiler und kann strukturelle Familienwohngebiete besser indizieren als der reine Anteil von Familienhaushalten, da Familien häufig in ihrer Gründungsphase noch umziehen. Die erklärte Gesamtvarianz beträgt 71,7 Prozent.

Schluss

Mit dem beschriebenen Verfahren ist ein valider, belastbarer Sozialindex entwickelt worden, der nun der Sozialplanung und der Beobachtung der sozialen Lage in den Freiburger Bezirken zur Verfügung steht.

Er ist sowohl im Informationsportal FR.ITZ³ als tabellarische Darstellung zu finden als auch im interaktiven Indikatorenatlas⁴. Dort sind jeweils die Indexwerte, deren Kategorisierungen als auch die zugrunde liegenden Indikatoren als Zeitreihe verfügbar. Auch die übrigen drei Indizes sind dort zu finden und können detailliert ausgewertet werden.

1 Backhaus, Klaus et al. (2021): Multivariate Analysemethoden, eine anwendungsorientierte Einführung, 16. Auflage, Wiesbaden: Springer Gabler.
 2 Tatsächlich gab es einen Vorgänger-Index, der allerdings nicht rein wissenschaftlich entwickelt wurde. Der Wunsch des Fachamtes war, diesen zu aktualisieren – die Autor*innen haben sich jedoch für die beschriebene Neuentwicklung entschieden.
 3 https://fritz.freiburg.de/asw/asw.dll?aw=Indikatoren/Indizes_Z_interaktiv
 4 <https://fritz.freiburg.de/indikatorenatlas/produktiv/atlas.html>

Abbildung 5: Komponentenladung, Kommunalitäten und Anteil nicht erklärter Varianz der ausgewählten Indikatoren für den Wohnungsneubauindex

	Komponentenladung	Kommunalitäten	Anteil Varianz, der nicht durch Faktor erklärt werden kann
Bevölkerungsentwicklung insgesamt (Veränderungsrate)	0,92	0,85	0,15
Anteil Neubauwohnungen letzte 5 Jahre	0,88	0,77	0,23
Veränderungsrate Wohnungsbestand Fertigstellungen	0,90	0,81	0,19
Veränderungsrate Haushalte	0,94	0,89	0,11

Abbildung 6: Komponentenladung, Kommunalitäten und Anteil nicht erklärter Varianz der ausgewählten Indikatoren für den Familienwohnindex

	Komponentenladung	Kommunalitäten	Anteil Varianz, der nicht durch Faktor erklärt werden kann
Jugendquotient	0,85	0,72	0,28
Durchschnittliche Anzahl Räume je Wohnung	0,71	0,51	0,49
Durchschnittliche Haushaltsgröße	0,94	0,89	0,11
Anteil Familienhaushalte	0,91	0,83	0,17
Pkw pro Haushalt	0,80	0,63	0,37