Das Bildungsbüro.

MAINZ · BINGEN
Landkreis

BERECHNUNGSVERFAHREN ZUR BILDUNG VON SOZIALINDIZES

mit besonderem Fokus auf faktorenanalytischen Verfahren

Jaqueline Brossart
Bildungsbüro Kreisverwaltung Mainz-Bingen
Georg-Rückert-Straße 11
55218 Ingelheim
brossart.jaqueline@mainz-bingen.de
06132 – 787 33 23

INHALT

1. Definition und Nutzen von (Sozial-)Indizes
2. Indexbildung
   2.1 Vorarbeit
   2.2 Berechnungsverfahren: Kategorisierung, Mittelwerts-/Summenindizes
3. Wiederholung
   3.1 (Empirische) Varianz
   3.2 z-Standardisierung
   3.3 Korrelation (nach Pearson)

INHALT (2)

4. Faktorenanalyse – was ist das und wofür?
   4.1 Formen
   4.2 Korrelationen
   4.3 Durchführung
   4.4 Interpretation Output
   4.5 Rotationsverfahren
   4.6 ... und dann?
5. Quellen

1. DEFINITION UND NUTZEN VON (SOZIAL-)INDIZES

- Ein Index bildet verschiedene Variablen gebündelt ab und erhöht so die Übersichtlichkeit
- Als „verdichtete Information“ ermöglicht er direkte Vergleiche verschiedener Einheiten (z. B. Individuen, Institutionen, Ortseinheiten). Hierdurch werden Ungleichheiten zwischen den Einheiten deutlich
- Kann vielfältig eingesetzt werden, z. B. zur Beobachtung von Entwicklungen über Zeit, gezielten Steuerung qualitativer Maßnahmen (z. B. Einrichten von Beratungsstellen in stark belasteten Sozialräumen) oder Verteilung von Ressourcen (Geld, Schulsozialarbeitsstellen)

2.1 INDEXBILDUNG – VORARBEIT

- Welches Konstrukt möchte ich messen? Z. B. Bildungsungleichheit, Umweltbewusstsein, Konservatismus, ...
- Welche Variablen beeinflussen das Konstrukt (z. B. Arbeitslosigkeit, Wahlbeteiligung, Migrationshintergrund, politische Parteipräferenz, ...)?
- Auf welcher Ebene betrachte ich mein Konstrukt (z. B. individuell, institutionell, kommunal, ...)?
- Welche Daten und welche Programme (z. B. Excel, Stata, SPSS, R, ..) stehen mir zur Verfügung?
- Wie möchte ich meinen Index rechnerisch bilden?

2.2 INDEXBILDUNG – BERECHNUNGSVERFAHREN: KATEGORISIERUNG, MITTELWERTS-/SUMMENINDIZES

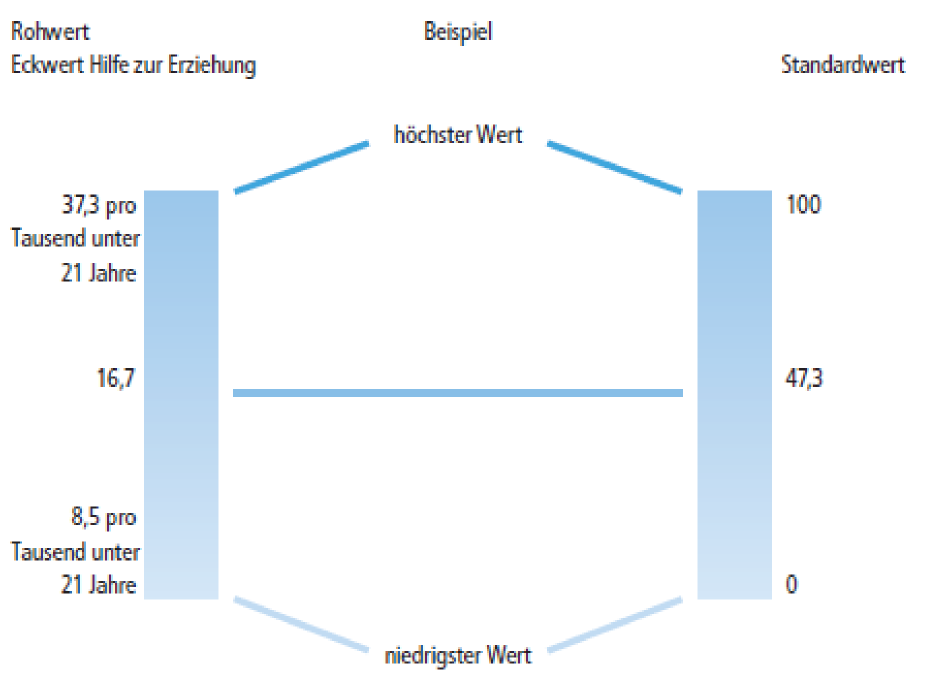
- Möglichkeit 1: Kategorisierung: Überführung der Einzelwerte in Kategorien und auszählen, z. B. „Institution xy liegt in 3 von 5 Kategorien im niedrigen Bereich“
- Möglichkeit 2: Scores (z-Standardisierung erforderlich!)
    - 2.1 Summenscore: Addition der Ausprägungen (z-Werte!) der Variablen
    - 2.2 Mittelwertsscore: Addition der Ausprägungen (z-Werte!), teilen durch die Anzahl der gültigen Variablenwerte
- Standardisierung der Ausprägungen, damit unterschiedlich hohe Werte vergleichbar sind (z. B. Arbeitslosenquote und Anteil der Abiturienten)
- Beispiel Rheingau-Taunus-Kreis 2012

BEISPIEL RHEINGAU-TAUNUS-KREIS 2012 (S.8-9)

V1 = Quote der Kinder in Bedarfsgemeinschaften (FL1 = 0,9)
V2 = Anteil der SGB II-Fälle an potentiell erwerbstätiger Bevölkerung (FL2 = 0,9)
V3 = Quote der nichtdeutschen Bevölkerung (FL3 = 0,7)
V4 = Quote Jugendhilfefälle (FL4 = 0,6)

- Z-Transformation der Werte
- Gewichtung der Werte (anhand der Faktorladungen)
- Transformation der Rohwerte auf eine Skala von 0-100
- Addition der Werte

$$S_{RTK} = V1*FL1+V2*FL2+V3*FL3+V4*FL4$$



Das Bildungsbüro.

MAINZ·BINGEN Landkreis

3.2 WIEDERHOLUNG: Z-STANDARDISIERUNG

- Die Z-Standarisierung setzt den Mittelwert der Verteilung auf 0 und die Streuung/Varianz auf 1, sodass negative Ausprägungen unter- und positive Ausprägungen überdurchschnittlich sind
- Die Ausprägungen der Variablen heißen nun „z-Werte“
- Wichtig: Die Z-Standardisierung ändert nicht die Verteilung, sondern staucht/streckt die Verteilung lediglich

3.2 WIEDERHOLUNG: Z-STANDARDISIERUNG (2)

- Auto 1: $x_1 = 4{,}8$l und $x_2 = 5{,}3$ Liter; Durchschnitt $(\bar{x})$ = 5,05l
- E-Auto 2: $x_1 = 12$ kWh und $x_2 = 18$ kWh ; Durchschnitt $(\bar{x})$ = 15 kWh
  - $s_{A1}^2 = 0{,}0625\ l^2 \Rightarrow s = 0{,}25$l
  - $s_{E2}^2 = 18{,}5\ kWh^2 \Rightarrow s = 4{,}301\ kWh$
- $z_i = \frac{x_i - \bar{x}}{s_i}$
  - $z_{A11} = \frac{4{,}8l - 5{,}05\ l}{0{,}25\ l} = -1 \Rightarrow x_1 =$
    $4{,}8l\ entspricht\ einer\ Standardabweichung\ vom\ Mittelwert\ von\ -1.$
  - $z_{E21} = \frac{12\ kWh - 15\ kWh}{4{,}301\ kWh} = -0.7$

3.1WIEDERHOLUNG – (EMPIRISCHE) VARIANZ

- Die Varianz (s^2) ist das Quadrat der Standardabweichung (s) einer Verteilung
- Die Standardabweichung misst die Streuung der Daten um den Mittelwert.
- $$\boldsymbol{s^2 = \frac{1}{n}\left[(x_1 - \overline{x})^2 + .. + (x_n - \overline{x})^2\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
- Beispiel (n=2):

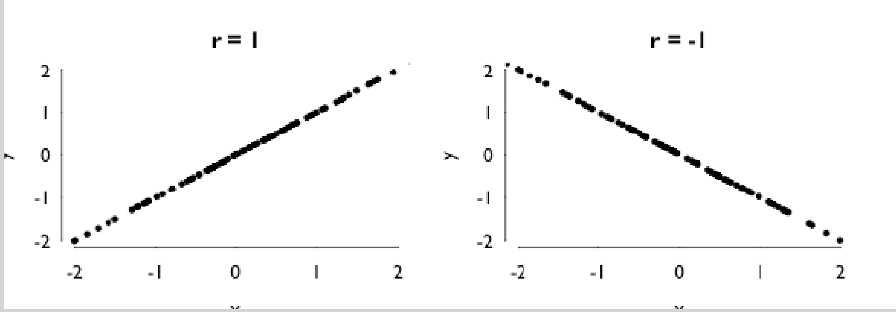
  Auto 1: $x_1 = 4{,}8$l und $x_2 = 5{,}3$ Liter; Durchschnitt $(\bar{x})$ = 5,05l

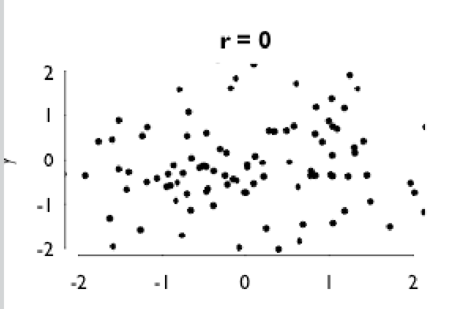
  - $s_{A1}^2 = \frac{1}{2}[(4{,}8l - 5{,}05l)^2 + (5{,}3l - 5{,}05l)^2 = \frac{1}{2}(0.0625l^2 + 0.0625l^2) =$ 0,0625l²
- Auto 2: $x_1 = 4{,}1l, x_2 = 5{,}5l; \bar{x} = 4{,}8l$
  - $\Rightarrow s_{A2}^2 = \frac{1}{2}[(4{,}1l - 4{,}8l)^2 + (5{,}5l - 4{,}8l)^2 = 0{,}49\ l^2$

3.3 WIEDERHOLUNG – KORRELATION (NACH PEARSON)

- Mit „Korrelation“ bezeichnet man den (linearen) Zusammenhang zwischen zwei Variablen
- Mithilfe des Korrelationskoeffizienten nach Pearson lässt sich die Stärke des Zusammenhangs einschätzen, er ist eine „Quasi-Zusammenfassung“ eines Streudiagramms und hat einen Wertebereich von -1 bis 1
- Leitfrage: Wie gut lassen sich die Werte des Streudiagramms auf einer gemeinsamen Linie abbilden?
- $r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} * \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{cov(x,y)}{s_x * s_y}$

3.4 WIEDERHOLUNG – KORRELATION (NACH PEARSON) (2)





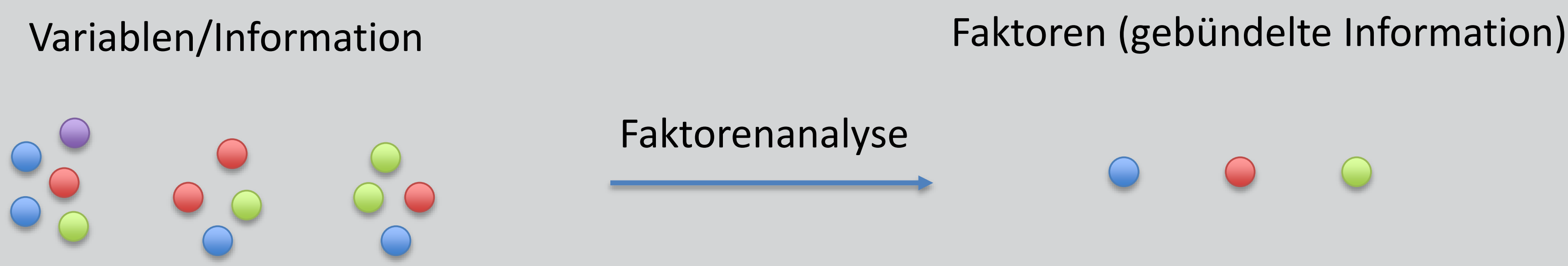
$r = 1$	perfekte positive Korrelation
$0 < r < 1$	positive Korrelation
$r = 0$	keine Korrelation
$-1 < r < 0$	negative Korrelation
$r = -1$	perfekte negative Korrelation

4. FAKTORENANALYSE – WAS IST DAS UND WOFÜR?

Ziel: Bündelung der gemeinsamen Informationen verschiedener Variablen in neuen Variablen (Faktoren), die den ursprünglichen Informationsgehalt möglichst gut wiederspiegeln sollen

Ausgangshypothese: Variablen messen teilweise ähnliche Dinge → Herausfiltern der Ähnlichkeiten durch Bildung neuer, „kompakterer“ Variablen (Faktoren)

Zu beachten: Sparsamkeitsprinzip vs. Informationsverlust → Abwägungssache



Das Bildungsbüro.

MAINZ · BINGEN
Landkreis

4. FAKTORENANALYSE – WAS IST DAS UND WOFÜR?

- Grundannahme: Wert einer Variable lässt sich additiv in eine gewichtete Summe aus Faktoren zerlegen
- $\boldsymbol{x_{im}} = \varepsilon_{i1} * \gamma_{m1} + \varepsilon_{i2} * \gamma_{m2} + .. + \varepsilon_{if} * \gamma_{mf} + \vartheta_{mi} = \boldsymbol{\sum_{j=1}^{f} \varepsilon_{i1} * \gamma_{mj} + \vartheta_{mi}}$
  - x_{im}: Beobachteter Wert der Person i auf der Variablen m
  - ε_{i1}: Wert der Person i auf dem Faktor j
  - γ_{ij}: Faktorladung der beobachteten Variable m auf Faktor j
  - f: Anzahl der dem Wert x_{im} zugrunde liegenden Faktoren
  - ϑ_{mi}: Residualvariable/Einzelrestvarianz

4.1 FAKTORENANALYSE – FORMEN

- Häufig verwandt: Konfirmatorische und Explorative Faktorenanalyse
  - Konfirmatorische Faktorenanalyse: Überprüfung konkreter Hypothesen zu den Faktoren und dem Faktorenmuster
  - Explorative Faktorenanalyse: „Unvoreingenommenes Untersuchen“, welche gemeinsamen Strukturen hinter den Variablen zu finden sind → strukturentdeckendes Verfahren, benötigt Interpretation durch den Forschenden

4.2 FAKTORENANALYSE: KORRELATIONEN UND GEMEINSAME VARIANZ

- Eignung der Daten überprüfen: Interkorrelationsmatrix
- Kaiser-Meyer-Olkin Maß (KMO): Wie viel gemeinsame Varianz beinhaltet die Interkorrelationsmatrix?
- Je weniger gemeinsame Varianz in der Interkorrelationsmatrix enthalten ist, desto kleiner wird der KMO!
- $KMO = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij*z}^2} \quad (\mathrm{i} \neq j)$
- $r_{ij} = Korrelationskoeffizient\ der\ Variablen\ i\ und\ j;\ r_{ij*z} =$ $partieller\ Korrelationskoeffizient\ nach\ Herauspartialisierung\ aller\ anderen\ Variablen$

4.2 FAKTORENANALYSE – KORRELATIONEN UND GEMEINSAME VARIANZ

KMO-Koeffizient	Eignung der Daten
.90	Sehr gut
.80 - .90	Gut
.70 - .79	Mittel
.60 - .69	Mäßig
.50 - .59	Schlecht
.50	Inkompatibel

Interpretation nach Kaiser und Rice 1974

- Für Eignungsprüfung einzelner Variablen: MSA (Measure of Sampling Adequacy), Interpretation identisch zum KMO-Maß
- Ggf. Variablen ausschließen

4.3 FAKTORENANALYSE – DURCHFÜHRUNG

- Verschiedene Verfahren: z. B. Hauptkomponentenanalyse, Hauptachsenanalyse, (Maximum-Likelihood-Faktorenanalyse)
  - Hauptkomponentenanalyse: Kommunalität wird mit 1 angenommen → keine Fehler- und/oder Residualvarianz
  - Hauptachsenanalyse: Kommunalität wird geschätzt
- Kommunalität (Wertebereich: 0 – 1) : Welcher Anteil der Varianz einer Variable kann durch alle Faktoren gemeinsam abgebildet werden? Wie viel Information der Variablen ist erhalten geblieben?
- Hier: Hauptachsenanalyse

4.4 FAKTORENANALYE – INTERPRETATION OUTPUT

- Programm bildet nach Faktorenanalysenbefehl iterativ Faktoren
- Ein Faktor macht nur dann Sinn, wenn er mehr Varianz als eine Variable erklärt => der Eigenwert sollte größer 1 sein, andernfalls sollte der Faktor nicht berücksichtigt werden
- Eigenwert (Eigenvalue): Gibt an, wie viel Varianz ein Faktor erklärt. Ein Eigenwert von 1 bedeutet, dass ein Faktor genauso viel Varianz wie eine Variable aufklärt
- Optional: Grafische Überprüfung durch Scree-Plot (Ellbogenkriterium): Faktoren bis „zum Knick“ können in die Analyse einfließen
- Faktorladungsmatrix: Gibt die Faktorladungen (Korrelationen) der Variablen auf die Faktoren wieder, Daumenregel: Faktorladung sollte größer als +/- 0.5 sein

4.4 FAKTORENANALYSE – INTERPRETATION OUTPUT (2)

- Angestrebt wird eine **Einfachsstruktur**. Diese ist gegeben, wenn eine Variable auf einen Faktor hoch auf alle anderen niedrig lädt
- Uniqueness: Anteil der Varianz, der nicht durch die gemeinsamen Faktoren erklärt werden kann; Gründe: Fehlervarianz, fehlende erklärende Variablen
- Uniqueness = 1 – Kommunalität ➔ Die Kommunalität sollte möglichst hoch, die Uniqueness möglichst klein sein!
- Anwendung eines Rotationsverfahren zur besseren Interpretierbarkeit (Einfachsstruktur wird angestrebt)

4.5 FAKTORENANALAYSE - ROTATIONSVERFAHREN

- Mathematisch betrachtet liegen Faktoren im n-dimensionalen (n = Anzahl der Faktoren) Datenraum. Die Korrelationen zwischen den Faktoren sind über die Winkel zueinander repräsentiert.
  - Unrotierte Lösung: Die Faktoren werden nach der Schätzung nicht nochmal neu im Raum angeordnet
  - Rotierte Lösung: Vektoren werden neu durch den geometrischen Raum gelegt (rotiert), sodass sich die Abstände zwischen den Vektoren verändern
    - ✓ Orthogonale Rotation: Die Vektoren stehen fix im 90°-Winkel zueinander
    - ✓ Oblique Rotation: Faktoren korrelieren untereinander

4.5 FAKTORENANANYLSE - ROTATIONSVERFAHREN (2)

- Die Entscheidung für das Rotationsverfahren kann inhaltlich getroffen werden, Leitfrage: Möchte ich, dass meine Faktoren untereinander korrelieren können, oder nicht?
- Kein Korrelieren: Orthogonale Rotation, z. B. sinnvoll, wenn die betrachteten Konstrukte sehr unterschiedlich sind oder eine Korrelation unerwünscht ist
- Korrelieren: Oblique Rotation, z. B. sinnvoll, wenn die Konstrukte teilweise ähnliche Sachverhalte beschreiben (z. B. Bildungsungleichheit und Demographieindex)

4.6 – FAKTORENANALYSE - … UND DANN?

- Ausschluss der ungeeigneten Variablen
- Erneute Durchführung: Korrelationsmatrix, Hauptachsenanalyse, Rotationsverfahren, Variablenausschluss, ...
- Wichtig: Wann eine Analyse beendet wird liegt im Ermessen des Forschenden!
- Im Anschluss z. B.
  - Verwendung der Variablen für einen Mittelwerts-/Summenindex
  - Weiterrechnen mit den gebildeten Faktoren
  - …

5. QUELLEN

5.1 Summen-/Mittelwertsindizes:

- Knüttel, Katharina 2019: Indexkonstruktion für ein sozialräumliches Bildungsmonitoring (Vortrag KOSMO-Fachgruppe C)
- Knüttel, Katharina; Kersting, Volker; Jehles, Nora 2019: Frühe Bildung trifft Armut? Das regionale Verhältnis von frühkindlicher Bildung und Kinderarmut in NRW: Analysen und Konzepte 01/2019; Gütersloh, online abrufbar
- Rheingau-Taunus-Kreis 2012: Sozialindex Rhein-Taunus-Kreis 2012; Bad Schwalbach, online abrufbar (beinhaltet sowohl Mittelwertsindex als auch Faktorenanalyse)

5. QUELLEN

- Kersting et al. 2009: Die A 40 – Der „Sozialäquator“ des Ruhrgebiets, in: Prossek, Achim et al.: Atlas der Metropole Ruhr: Vielfalt und Wandel des Ruhrgebietes im Kartenbild. Köln, S. 142 – 145, online abrufbar
- Stadt Gelsenkirchen 2018: Gesellschaftliche Teilhabechancen von Gelsenkirchner Kindern – Entwicklung und Stand 2018: Grundlage für eine sozialräumliche Strategieentwicklung, online abrufbar

5. QUELLEN

5.2 Faktorenanalysen:

- Stadt Hagen 2017: Hagener Sozialraumindex 2015; Stadt Hagen (online abrufbar)
- Emanuel Hartkopf 2006: Sozialräumliche Strukturen und Disparitäten in Bochum: Zusammenfassung einer faktorialökologischen Untersuchung der aktuellen demografischen und sozio-ökonomischen Situation auf Ortsteilebene; Bochum (online abrufbar)
- Groos, Thomas 2014: Schulsozialindices für die Grundschulen in Mühlheim an der Ruhr – Aktualisierung der Indices für die Grundschulen und Entwicklung von Wohnumfeldprofilen; Bochum (online abrufbar)

5. QUELLEN

5.3 Statistisches:

- Eric Klopp o. J.: Explorative Faktorenanalysen, o. O.; abrufbar unter https://eric-klopp.de/texte/explorative-faktorenanalyse.php
- Wolff, Hans Georg; Bacher, Johann 2010: Hauptkomponentenanalyse und explorative Faktorenanalyse, in: C. Wolf; H. Best (Hrsg.) 2010: Handbuch der sozialwissenschaftlichen Datenanalyse; VS Verlag für Sozialwissenschaften | Springer Fachmedien Wiesbaden gmbH 2010
- Müller, Michael 2016: Stata 13: Explorative Faktorenanalyse, abrufbar unter https://www.youtube.com/watch?v=HAc9c47B5tE