

6 Typenbildung und Repertoireforschung

Das zentrale Ergebnis der Befragung soll die Bildung von Typen von Informationsrepertoires sein. Typologien sind häufig ein Schritt zu erklärenden Theorien, eine Art „Bindeglied zwischen empirischer Beobachtung und Theoriebildung“ (Reith & Kelle, 2017, S. 571). Mit dieser Typologie von Informationsrepertoires soll das Zusammenspiel von Beeinträchtigung, Kontext- und personenbezogenen Faktoren und den Eigenschaften von Medien analysiert werden, das förderlich oder hinderlich für die Teilhabe an öffentlicher Kommunikation von Menschen mit Beeinträchtigungen wirken kann. Wie in Kap.4.3 und 5 begründet, werden die Cluster ausschließlich auf Grundlage der Frage nach den Nutzungshäufigkeiten von Informationsquellen gebildet und anschließend Unterschiede in den Kontextfaktoren zwischen den Clustern verglichen.

6.1 Clusteranalyse

Die Clusteranalyse gehört zu den explorativen Verfahren der multivariaten statistischen Analyseverfahren, mit denen sich Strukturen in empirischen Beobachtungen entdecken lassen. Mithilfe der Clusteranalyse werden Objekte nach Ähnlichkeiten in den Ausprägungen ausgewählter Merkmale zu homogenen Gruppen zusammengefasst. Clusteranalysen sind in den Sozialwissenschaften eine gängige Methode, um Typen zu bilden wie zum Beispiel Lebens- und Konsumstile oder Werteorientierungstypen (Bacher, Pöge & Wenzig, 2010, S. 7; Schendera, 2010, S. 2). Die Sinus-Milieus sind ein prominentes Beispiel für eine solche Typenbildung, von 2005 bis 2016 hat die Studie Massenkommunikation die Mediennutzung der Sinus-Milieus untersucht (Krupp & Breunig, 2015, S. 190–207).

Es gibt unterschiedliche Verfahren bei der Clusteranalyse, das Grundprinzip ist bei allen Verfahren gleich. In den Clustern sind die Merkmals-träger oder Objekte bezüglich der festgelegten Merkmale möglichst homogen (hohe Intracluster-Homogenität), während sie sich zu den Merkmals-trägern der anderen Gruppen bezüglich der Merkmale möglichst stark unterscheiden (geringe Intercluster-Homogenität) (Bacher et al., 2010, S. 16; Baur & Blasius, 2019, S. 1392; Eckstein, 2016, S. 428; Schendera, 2010, S. 8). Die Clusteranalyse beruht auf der Annahme, dass es im Daten-

satz „natürliche Klassen von Merkmalsträgern“ gibt, die sich bezüglich bestimmter Merkmale – in dieser Arbeit die Nutzung von Informationsquellen – ähnlich sind und sich gleichzeitig von anderen Klassen unterscheiden (Fromm, 2012, S. 191). Die Clusteranalyse ist ein „objektiv klassifizierendes Verfahren“, es ist selbst kein sinnstiftendes Verfahren (Schendera, 2010, S. 7). Es bedarf einer theoretischen Begründung, ob die Clusteranalyse sinnvolle Strukturen aufdeckt. Die in den Kapiteln zum Stand der Forschung und zu den Informationsrepertoires (Kap. 3 und 4) vorgestellten Studien sprechen für die Annahme, dass es bei der Zusammenstellung von Informationsquellen Muster gibt, nach denen Menschen die Quellen kombinieren. Ziel dieses Forschungsvorhabens ist es zu ermitteln, inwieweit sich die Cluster auch in der Wechselwirkung von Funktionsbeeinträchtigungen mit den Kontextfaktoren des ICF-Analysesystems unterscheiden.

Nach Bacher et al. werden folgende Anforderungen an Clusteranalysen gestellt (Bacher et al., 2010, S. 18):

- 1 Die Objekte innerhalb eines Clusters sollen sich in Bezug auf die festgelegten Merkmale ähnlich sein.
- 2 Die Cluster sollen sich voneinander unterscheiden.
- 3 Die Clusterlösung soll in der Lage sein, die Variation in den Daten zu erklären.
- 4 Die Clusterlösung soll auch bei geringfügigen Änderungen von Daten und bei Anwendung anderer Clusterverfahren stabil bleiben.
- 5 „Die Cluster sollen inhaltlich gut interpretierbar sein. Den Clustern sollen inhaltlich sinnvolle Namen gegeben werden können. Im Idealfall sollen die Namen aus einer Theorie abgeleitet werden.“
- 6 Die Cluster sollen (inhaltlich) valide sein. Die Cluster sollen mit externen Variablen korrelieren, von denen bekannt ist, dass sie im Zusammenhang mit den Typen stehen, die aber nicht in die Bildung der Cluster eingehen“ (Bacher et al., 2010, S. 18).
- 7 Die Anzahl der Cluster soll überschaubar sein, weil dies für die Kriterien 5 (Interpretierbarkeit) und 4 (Stabilität) förderlich ist.
- 8 Die Cluster sollen eine gewisse Mindestgröße haben, weil dies zur Stabilität beiträgt.

Die Punkte 7 und 8 werden nach Bacher et al. „nur mitunter“ für gute Clusterlösung gefordert (Schendera S. 18, 5 Punkte für eine gute Clusterlösung). Diese Kriterien machen bereits deutlich, dass die Clusteranalyse theoriegeleitet ist und erst mit einer entsprechenden Theorie einen Sinn ergibt.

Die folgenden Kriterien für eine gute Typologie von Kuckartz, die er unabhängig von der Art der Analyseverfahren aufstellt, verdeutlichen, wie Typenbildung und Theorie zusammenhängen (Kuckartz, 2010, S. 565):

- 1 Jedes Objekt wird nur einem Typ zugeordnet.
- 2 Die Typologie folgt dem „Prinzip der Sparsamkeit“, es sollte nur so wenige Typen/Cluster wie möglich und so viel wie nötig geben.
- 3 Zwischen den einzelnen Typen und dem Ganzen soll es einen Zusammenhang geben, sie beziehen sich aufeinander und weisen eine erkennbare Gestalt auf.
- 4 Die für die Typenbildung ausgewählten Merkmale sind relevant für die Fragestellung und begründet.
- 5 Die Merkmale und Dimensionen der Typenbildung werden offengelegt und der Merkmalsraum nachvollziehbar beschrieben.
- 6 „Die Typologie ist fruchtbar im Hinblick auf die Entdeckung neuer Phänomene und erweist sich in neuen Forschungsfeldern als heuristisch brauchbar“ (Kuckartz, 2010, S. 565)

Die von Bacher et al. aufgestellten Anforderungen an die Clusteranalyse entsprechen den Kriterien Kuckartz an eine gute Typologie.

6.1.1 Verfahren der Clusteranalyse

Wie oben bereits erwähnt, gibt es nicht die Clusteranalyse, sondern verschiedene Verfahren, die je nach Zweck und Datenmaterial ausgewählt werden. Die meisten Verfahren beruhen auf der Bestimmung von Ähnlichkeits- und Distanzmaßen. Die Abstände der Objekte voneinander werden in einem mehrdimensionalen Raum gemessen und dann mit unterschiedlichen Algorithmen zu Clustern geordnet. Einige Algorithmen berechnen die Zuordnung nicht auf der Grundlage von Ähnlichkeiten und Distanzen der Objekte, sondern von Clusterzentren.

Es gibt drei große Verfahrensgruppen von Clusteranalysen: unvollständige, auch geometrische Verfahren genannt, deterministische und probabilistische Clusteranalyseverfahren (Bacher et al., 2010, S. 18; Eckstein, 2016, S. 429). Bei geometrischen Verfahren werden die Clustereinteilungen aufgrund von räumlichen Darstellungen vorgenommen. Dies ist nur möglich, wenn drei und weniger Variablen in die Klassifikation einfließen, da die Lage der Klassifikationsobjekte nur in einem zwei- oder dreidimensionalen Raum darstellbar ist (Bacher et al., 2010, S. 19; Eckstein, 2016, S. 431). Bei deterministischen Verfahren werden die Objekte genau einem Cluster zugeordnet (Wahrscheinlichkeit von 1 oder 0), während probabilistische

Verfahren die Objekte mit einer dazwischenliegenden Wahrscheinlichkeit einem Cluster zuzuordnen (Bacher et al., 2010, S. 20).

Für das Ziel dieser Arbeit eignen sich lediglich die deterministischen Verfahren, da jede*r Befragte genau einem Cluster zugeordnet werden soll. Die deterministischen Verfahren werden wiederum in hierarchische und partitionierende Verfahren unterteilt. Hierarchische Verfahren gehen sukzessive vor und bilden die Cluster schrittweise. Dadurch entstehen Gruppierungen auf jeweils unterschiedlichen Distanz- und Ähnlichkeits-ebenen und das Ergebnis des Prozesses ist anschaulich in einer Hierarchie von Clustern darstellbar, etwa in Dendrogrammen. Man unterscheidet zwischen zwei Verfahrensgruppen, den divisiven und den agglomerativen Verfahren. Divisive Verfahren gehen von einem alle Objekte (oder Variablen) enthaltenen Cluster aus und teilen diese schrittweise in immer kleiner werdende Cluster auf. Die agglomerativen Verfahren gehen umgekehrt vor: Die einzelnen Elemente werden sukzessiv zu immer umfangreicheren Clustern zusammengefasst (Steinhausen & Langer, 1977, S. 73). Die agglomerativen Verfahren gelten als die bedeutendere Teilklasse der hierarchischen Verfahren (Schlittgen, 2009, S. 398). Deshalb werden die divisiven Verfahren hier nicht berücksichtigt.

Bei partitionierenden Verfahren muss die Anzahl der Cluster vorgegeben werden. Sie optimieren schrittweise die Zuordnung der Elemente zu den Clustern nach einem vorgegebenen Kriterium, so dass die Varianz in den Clustern minimiert wird (Bacher et al., 2010, S. 19). Das bekannteste partitionierende Clusterverfahren ist das k-mean-Verfahren. Die partitionierenden Verfahren sind rechenaufwendiger, da immer wieder Neuzuordnungen der Cluster stattfinden.

In der Praxis wird häufig mehrstufig vorgegangen und zunächst die Anzahl der Cluster über ein hierarchisches Verfahren bestimmt. Im zweiten Schritt werden die Cluster dann mit dem k-means-Verfahren optimiert (Janssen & Laatz, 2017, S. 502).

6.1.2 Hierarchisch-agglomerative Verfahren

Hierarchisch-agglomerative Verfahren fassen sukzessive ungruppierte Objekte zusammen, bis am Ende ein einziges Cluster entstehen würde. Es handelt sich also um ein iteratives Verfahren, bei dem zunächst ein Objekt mit dem nächstähnlichen Objekt in einer Ähnlichkeits- bzw. Distanzmatrix zu einem Cluster fusioniert wird. Eine neue Matrix wird berechnet und wiederum das nächst ähnliche Objekt gefunden und mit dem Cluster

fusioniert. Alle verschiedenen Verfahren dieser Art stimmen in dem allgemeinen Verfahren überein, die Algorithmen unterscheiden sich jedoch darin, wie sie die Ähnlichkeiten der bzw. Distanz zwischen den Clustern berechnen (Schendera, 2010, S. 24). Welches Verfahren man auswählt, hängt davon ab, welche Anforderung man an die Cluster aufgrund der Theorie stellt und wie das Messniveau der Variablen und Verteilungsmerkmale des Samples beschaffen sind.

Nach Abwägung der mit den Verfahren verbundenen Anforderungen und Voraussetzung wird das Ward-Verfahren für die Clusterbildung ausgewählt. Das Ward-Verfahren wird in der Literatur für objektorientierte Clusterverfahren empfohlen (Bacher et al., 2010, S. 295) und gilt als leistungsstarkes Verfahren. Es strebt homogene und annähernd gleich große Cluster an. In die Clusteranalyse fließen nur Variablen mit der gleichen Skala ein, sodass eine Standardisierung nicht nötig ist. Bei Clusterzentrenverfahren wie Ward und k-means müssten die Skalen andernfalls vorher standardisiert werden (Schendera, 2010, S. 14).

Kurze Beschreibung des Ward-Verfahrens

Die Cluster werden über den minimalen Abstand der Interclustervarianz fusioniert, die Clusterung erfolgt also nicht über die Abstandswerte der Objekte, sondern über ein Varianzkriterium: Die Gesamtsumme der internen Abstandsquadrate der Gruppenzentroide muss minimal sein.

„Es werden vielmehr Cluster (bzw. Cluster und Objekte) mit dem Ziel vereinigt, den Zuwachs für ein Maß der Heterogenität eines Clusters zu minimieren. Als Maß für die Heterogenität wird die Summe der quadrierten Euklidischen Distanzen (auch Fehlerquadratsumme genannt) der Objekte zum Zentrum des Clusters (Zentroid) gewählt. Auf jeder Stufe wird also das Clusterpaar fusioniert, das zum kleinsten Zuwachs der Fehlerquadratsumme im neuen Cluster führt“ (Janssen & Laatz, 2017, S. 501–502).

Die Matrix dieses Verfahrens enthält nicht die Distanzen zwischen Gruppen oder Elementen, sondern den jeweiligen Zuwachs an Heterogenität (Varianzkriterium) bei der Verschmelzung der Cluster bzw. Elemente (Steinhausen & Langer, 1977, S. 81). Das Ward-Verfahren strebt Cluster von annähernd gleicher Größe an, ist aber schlecht darin, „kleine oder langgezogene Gruppen“ zu erkennen und anfällig für Ausreißer (Schendera, 2010, S. 27). Es ist in der Praxis weit verbreitet (Backhaus, Erichson,

Plinke & Weiber, 2018, S. 465). Bacher et al. empfehlen es bei einer objektorientierten Analyse, d.h. Fälle und nicht Variablen werden geclustert, und bei kleineren Fallzahlen (Bacher et al., 2010, S. 295). Die Datenmenge von 617 Datensätzen erlaubt ein Ward-Verfahren. Untersuchungen haben ergeben, dass „das Ward-Verfahren im Vergleich zu anderen Algorithmen in den meisten Fällen sehr gute Partitionen findet und die Elemente „richtig“ den Gruppen zuordnet“ (Backhaus et al., 2018, S. 470). Voraussetzung sind, dass die Variablen auf metrischem Niveau sind und das Sample keine Ausreißer aufweist und etwa gleich große Gruppen mit ähnlicher Ausdehnung zu erwarten sind (Backhaus et al., 2018, S. 470).

Mit dem Wardverfahren wird die zu erwartende Anzahl der Cluster bestimmt und eine Clusterlösung errechnet. Um sicherzugehen, eine möglichst gute Lösung zu ermitteln, wird anschließend die Ward-Lösung mit der Clusterlösung verglichen, die mit dem k-means-Verfahren, einem partitionierenden Verfahren, erzielt wird. Dabei wird im k-means-Verfahren die Clusteranzahl vorgegeben, die mit dem Ward-Verfahren errechnet wurde. Bei dem Vergleich der Lösungen wird nach inhaltlicher Plausibilität entschieden, welche der beiden Lösungen schlussendlich ausgewählt wird.

Kurze Beschreibung des partitionierenden Verfahrens k-mean

Bei den partitionierenden Verfahren bleibt die Anzahl der Cluster gleich, aber die gebildeten Cluster können während der Rechenoperation aufgelöst und neu gebildet werden. Diese Verfahrensgruppe wird vor allem bei großen Datenmengen angewandt, bei denen man aus der Theorie heraus schon eine Vorstellung von der Clusteranzahl hat (Eckstein, 2016, S. 442).

Das k-means-Verfahren, auch Clusterzentrenanalyse genannt, eignet sich nur für metrische Daten (Janssen & Laatz, 2017, S. 502). Zunächst werden die Elemente des Samples zufällig in die vorgegebene Anzahl (k) von Clustern zugeordnet. Danach werden in iterativen Schritten die Cluster optimiert, indem die Objekte nach der kleinsten euklidischen Distanz zum Clusterzentrum zugeordnet werden. Wie beim Ward-Verfahren wird in den Clustern die Fehlerquadratsumme minimiert (Janssen & Laatz, 2017, S. 502). Die Clusterzentren werden bei jedem Schritt neu berechnet und die Objekte wiederum neu zugeordnet. Der Vorgang wird solange wiederholt, bis eine optimale Clusterzuordnung erreicht wird, das heißt, bis sich die Werte der Clusterzentren nicht mehr verändern bzw. unter einem gewissen Schwellenwert liegen, oder bis die maximale Anzahl der

vorgegebenen Rechenoperationen (Iterationen) überschritten ist (Schendera, 2010, S. 118).

6.2 Vorgeschaltete Faktorenanalyse

Der Clusteranalyse wird häufig eine Faktorenanalyse vorgeschaltet. Die Faktorenanalyse zählt ebenfalls zu den strukturentdeckenden Verfahren, „mit deren Hilfe Muster in Daten aufgedeckt werden, die nicht als Zusammenhänge von unabhängigen und abhängigen Variablen interpretiert werden können“ (Fromm, 2012, S. 59). Sie identifiziert Variablen, die hoch miteinander korrelieren und ordnet sie zu Faktoren (Backhaus et al., 2018, S. 366). Variablen aus unterschiedlichen Faktoren korrelieren wenig miteinander. Die Faktorenanalyse wird vor allem für drei Ziele eingesetzt:

- um latente Strukturen in Variablensets aufzudecken,
- um mehrteilige Messinstrumente zu entwickeln und zu prüfen,
- um Daten zu reduzieren.

Bei der Datenreduktion werden Messwerte von Variablen für weitere Analysen durch eine geringere Anzahl von Werten der dahinterstehenden Faktoren ersetzt (Janssen & Laatz, 2017, S. 577). Wenn die extrahierten Faktoren eine „sachlogisch plausibel zu benennende Klassifikation der empirisch beobachteten Variablen“ sind, können sie die Basis für weitere statistische Analysen bilden (Eckstein, 2016, S. 323). Am gebräuchlichsten ist das Verfahren der Hauptkomponentenanalyse (HKM) oder auch Principal Component Analysis (PCA) genannt.

Die Analyse erfolgt in drei Schritten: Es wird eine Korrelationsmatrix der einbezogenen Variablen (Pearsons) aufgestellt, aus der die Eigenwerte und Eigenvektoren der Faktoren berechnet werden (Faktorextraktion), alle Faktoren mit einem Eigenwert ab 1,0 werden extrahiert. Durch eine Faktorenrotation werden Faktoren besser interpretierbar. Die Faktorenladung wird für die Faktoren in einer Matrix ausgegeben, auf dieser Grundlage lassen sich die Faktoren interpretieren und schließlich werden die Variablen anhand ihrer Faktorwerte den Faktoren zugeordnet (Bühl, 2016, S. 599; Eckstein, 2016, S. 325).

Faktorenanalysen können für die Hypothesenbildung, wie Variablen zusammenhängen, „ausgesprochen inspirierend“ wirken, „d. h., es fällt meistens nicht schwer, diverse Hypothesen darüber zu generieren, was ein Faktor inhaltlich bedeutet bzw. was „hinter“ den Variablen eines Faktors steht“ (Döring & Bortz, 2016, S. 624). Die faktoranalytische Untersuchung der 31 Variablen zur Häufigkeit der Nutzung der Informationsquellen gibt

bereits wichtige Hinweise darauf, welche Informationsquellen von den Befragten kombiniert werden. Mithilfe der Analyse lässt sich ablesen, welche Medien eher konkordant genutzt werden (wer die eine Quelle häufiger nutzt, nutzt auch die andere häufiger) oder eher konkurrierend (wer die eine Quelle häufiger nutzt, nutzt die anderer seltener) (vgl. Hasebrink & Schmidt, 2012, S. 15). So wäre es zum Beispiel möglich, dass Informationsquellen der gleichen Mediengattung häufig kombiniert werden, weil sie den Bedarfen bei einer Sinnesbeeinträchtigung gut entsprechen, während crossmediale Kombinationen selten sind. Die Kombination von Informationsquellen kann auch durch die Dimensionen der Informationsbedürfnisse geprägt sein, da verschiedene Dimensionen von Informationsbedürfnissen besonders gut mit bestimmten Quellen befriedigt werden können, wie etwa gruppenbezogene Bedürfnisse mit sozialen Medien.

Faktorenanalysen werden Clusteranalysen vorgeschaltet, um hoch korrelierende Variablen auf wenige Faktoren zu verdichten, die nicht mehr miteinander korrelieren. Nach Schendera sind diese Faktoren brauchbarer für eine Clusteranalyse als viele hoch korrelierende Einzelfaktoren, da man damit „das Problem der Messfehlerbehaftetheit von Variablenwerten“ umgehe (Schendera, 2010, S. 19). Man geht dabei davon aus, dass Messfehler bei einzelnen Variablen zufällig und voneinander unabhängig auftreten. „Daher kann alle gemeinsame Varianz der beobachteten Variablen als messfehlerfreie, wahre Varianz des interessierenden Konstruktes aufgefasst werden“ (Döring & Bortz, 2016, S. 949).

Dafür ist entscheidend, dass die Faktoren sinnvoll sind, also mit der Theorie erklärbar, und dass bei der Ermittlung keine Informationen verloren gehen (Schendera, 2010, S. 19).

Auch Bacher et al. empfehlen die Bildung abgeleiteter Variablen für Clusteranalysen, da „Messfehler und somit der Einfluss irrelevanter Variablen reduziert [wird, d. Verf.], was insbesondere für die deterministische Clusteranalyse von Bedeutung ist“ (Bacher et al., 2010, S. 459).