Methodisches Vorgehen

8

8.1 Studien und Stichproben

Die vorliegende Arbeit nutzt Daten aus drei Forschungsprojekten und verbindet in der Untersuchung individueller Bedarfe und beruflicher Entwicklung querschnittliche und längsschnittliche Analysedesigns miteinander. Im Folgenden sollen die drei Forschungsprojekte mit den jeweiligen Zielen kurz vorgestellt werden. Tabelle 8.1 gibt einen Überblick über die vier in dieser Arbeit vorgestellten empirischen Studien, ihre Datengrundlagen sowie die jeweiligen Stichprobengrößen und -zusammensetzungen.

Die Berliner Berufs- und Studienorientierungsstudie (BeBest-Studie)

Die *Berliner Berufs- und Studienorientierungsstudie* (BeBest-Studie) wurde in den Jahren 2013 bis 2016 unter Leitung von Prof.ⁱⁿ Dr. Angela Ittel vom Fachgebiet Pädagogische Psychologie an der Technischen Universität Berlin im Auftrag der Senatsverwaltung für Bildung, Jugend und Wissenschaft Berlin durchgeführt.

Ziel der Studie war es, die Art und Durchführung von berufsorientierenden Maßnahmen an Berliner Integrierten Sekundarschulen (ISS) und Gymnasien eingehend zu erfassen und auszuwerten. Die Studie war in einem Mixed-Method-Design bestehend aus qualitativen Interviews, Fokusgruppen und quantitativen Befragungen konzipiert.

Anhand strukturierter Experteninterviews mit jeweils einem Mitglied der Schulleitung bzw. der in der Schule für die Koordination der beruflichen Orientierung verantwortlichen Person wurden die schulinternen Konzepte beruflicher Orientierung der 26 teilnehmenden Schulen erhoben. Die Perspektive der Lehrkräfte auf Chancen und Herausforderungen hinsichtlich der schulinternen Umsetzung beruflicher Orientierung wurden in fünf Fokusgruppengesprächen erfragt. Der Stand

Tabelle 8.1 Empirische Studien

Studie	Datenquelle	<i>N</i>	Alter in Jahren
Studie 1: Zusammenhänge von Berufsorientierungsmaßnahmen und persönlichen Merkmalen Jugendlicher als Determinanten individualisierter schulischer Berufsorientierung	• BeBest-Studie (MZIP 2)	1698	10–18 <i>M</i> = 14.75, <i>SD</i> = 1.11
Studie 2: Normwerte der Berufswahlkompetenz: Eine diagnostische Chance zur individuellen Förderung?	• BeBest-Studie (MZIP 3) • Pilotstudie Potentialanalyse (MZIP 3) • BK-Inno (MZIP 1)	1065	13–57 <i>M</i> = 16.70, <i>SD</i> = 2.28
Studie 3: Latente Berufswahlkompetenzprofile an weiterführenden allgemeinbildenden Schulen	• Pilotstudie Potentialanalyse (MZIP 3)	303	15–18 <i>M</i> = 15.94, <i>SD</i> = 0.48
Studie 4: Berufswahlkompetenzentwicklung an weiterführenden allgemeinbildenden Schulen: Eine latente Transitionsanalyse	• Pilotstudie Potentialanalyse (MZIP 1–3)	<i>N</i> _{MZIP1} = 309 <i>N</i> _{MZIP2} = 337 <i>N</i> _{MZIP3} = 303 <i>N</i> _{MZIP1–3} = 220	MZIP 1: 13–16 <i>M</i> = 14.82, <i>SD</i> = 0.50

Anmerkung. *N* = Anzahl, MZIP = Messzeitpunkt, *M* = Mittelwert, *SD* = Standardabweichung

der beruflichen Entwicklung der Schüler*innen ab der Jahrgangsstufe sieben, ihre Teilnahme an verschiedenen beruflichen Orientierungsmaßnahmen sowie ihre Wahrnehmung der besuchten Maßnahmen wurde zu zwei Messzeitpunkten mittels Fragebögen ermittelt. Zum ersten Messzeitpunkt, der im Dezember 2013 und Januar 2014 stattfand, nahmen insgesamt 3037 Jugendliche (46 % weiblich) der Jahrgangsstufen 7–13 an 14 ISS ohne gymnasiale Oberstufe, 2 ISS mit gymnasialer Oberstufe und an 12 Gymnasien an der Befragung teil. An der zweiten Erhebung im Herbst 2014 nahmen 3.599 Jugendliche (46 % weiblich) der bereits zum ersten Messzeitpunkt befragten Klassen teil. Eine dritte Befragung zur Vertiefung der vorherigen Ergebnisse wurde im Juli 2016 mit einer Teilstichprobe bestehend aus 324 Schüler*innen (48 % weiblich) der Jahrgangsstufen neun und zehn an vier

Schulen durchgeführt. Aus den Ergebnissen der Teilprojekte wurden übergreifende Handlungsempfehlungen formuliert. Eine ausführliche Beschreibung der Studienergebnisse und zentralen Handlungsempfehlungen finden sich bei Ohlemann et al. (2016). Wie der Tabelle 8.1 zu entnehmen ist, bilden die im Rahmen der BeBest-Studie erfassten Daten die Basis der in Kapitel 9 und 10 vorgestellten Studien. Die Daten von Messzeitpunkt zwei der Jahrgänge 8–10 wurden für die Analyse der Zusammenhänge von berufsorientierenden Maßnahmen und persönlichen Merkmalen Jugendlicher in Kapitel 9 herangezogen. Die Daten des dritten Messzeitpunkts der Jahrgänge 9–10 flossen in die Auswertungen der in Kapitel 10 vorgestellten Studie zur Überprüfung eines Kernwerts der Berufswahlkompetenz ein.

Innovationen in der Ausbildung zum Lehramt an Berufskollegs am Studienstandort Münster (BK-Inno)

Das Vorhaben *Innovationen in der Ausbildung zum Lehramt für Berufskollegs am Studienstandort Münster* (BK-Inno) wurde in den Jahren 2014 bis 2018 als Verbundprojekt durch das Institut für Berufliche Lehrerbildung an der Fachhochschule Münster und das Institut für Erziehungswissenschaft an der Westfälischen Wilhelms-Universität Münster (WWU) durchgeführt. An der WWU Münster konnten im Rahmen des Teilprojekts „Biographische Berufsbildungsplanung und Berufsorientierung“ Erkenntnisse sowohl über berufliche Entwicklungsprozesse von Schüler*innen am Berufskolleg als auch über Berufswahlmotive Studierender im beruflichen Lehramt gewonnen werden (Diesel-Lange, Morgenstern & Keune, 2017). Das nordrhein-westfälische Ministerium für Schule und Weiterbildung sowie das Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen förderten die Studie.

Die Daten von 454 Teilnehmenden (48 % weiblich) aus verschiedenen Ausbildungsgängen an insgesamt vier Berufsschulen in Nordrhein-Westfalen ergeben eine Teilstichprobe der in Kapitel 10 vorgestellten Studie (siehe Tabelle 8.1).

Das Pilotprojekt Potentialanalyse

Das Pilotprojekt *Potentialanalyse* wurde – gefördert von der Sparkasse Münsterland-Ost – in den Jahren 2014 bis 2016 in Kooperation der WWU Münster, der Friedrich-Schiller-Universität Jena und der Fachhochschule des Mittelstands Bielefeld durchgeführt. Ziel der Studie war es, die berufliche Entwicklung Jugendlicher in Schulformen, die zur Hochschulreife führen, zu beleuchten und den Effekt der Teilnahme an der berufsorientierenden Maßnahme Potentialanalyse zu untersuchen. Die Untersuchung war als längsschnittliche Pre-Post-Studie im Experimental-Kontrollgruppendesign mit vier Messzeitpunkten angelegt.

Beginnend im Schuljahr 2013/14 wurden 309 Schüler*innen (56 % weiblich) der Jahrgangsstufe neun an drei weiterführenden Schulen im Kreis Warendorf ein erstes Mal befragt. Diese erste Befragung fand vor der Teilnahme an der Potentialanalyse statt. Insgesamt neun Klassen an zwei Gymnasien sowie der teilnehmenden Gesamtschule stellten die Kontrollgruppe dar. Insofern nahmen sie nicht an der Potentialanalyse teil, die als eintägiges Planspiel konzipiert war. Der zweite Messzeitpunkt fand direkt im Anschluss an die Durchführung der Potentialanalyse statt. Die dritte und vierte Erhebung folgte jeweils mit einem Jahr Abstand im Schuljahr 2014/15 bzw. 2015/16. Wie aus der Tabelle 8.1 ersichtlich wird, basieren die empirischen Studien in Kapitel 10 und 11 unter anderem auf den Daten des dritten Messzeitpunkts. Die Untersuchung latenter Transitionsprofile in Kapitel 12 basiert auf den längsschnittlichen Daten der Messzeitpunkte eins bis drei dieser Studie. Aufgrund der Komplexität des Analyseverfahrens, das im folgenden Kapitel im Detail erläutert werden wird, wurde davon abgesehen, den vierten Messzeitpunkt in diese Analyse zu integrieren.

8.2 Statistische Auswertungsverfahren

In den empirischen Studien der Kapitel 9 bis 12 wurden verschiedene statistische Verfahren zur Beantwortung der jeweiligen Fragestellung herangezogen. Diese statistischen Auswertungsmethoden sollen hier entsprechend der Reihenfolge ihrer Verwendung nach vorgestellt und von anderen Auswertungsmethoden abgegrenzt werden.

Zusammenhangsanalysen in Kapitel 9: multiple lineare Regressionen

Zur Untersuchung von Zusammenhängen zwischen verschiedenen Variablen eignen sich beispielsweise Korrelations- oder Regressionsanalysen. In der in Kapitel 9 vorgestellten empirischen Studie wird der Zusammenhang zwischen berufsorientierenden Maßnahmen, den persönlichen Merkmalen der teilnehmenden Jugendlichen und deren subjektiv wahrgenommener Stand der beruflichen Orientierung betrachtet. Im Folgenden wird das verwendete Verfahren multipler linearer Regressionen unter Einbezug von Mehrfachinteraktionen dargestellt und erläutert.

Ziel und Prinzip der linearen Regression

Ziel von Regressionsanalysen ist es, Veränderungen in einer abhängigen Variablen, dem Kriterium oder auch Regressanden genannt, anhand von Veränderungen einer oder mehrerer unabhängiger Variablen, den Prädiktoren oder Regressoren, zu

schätzen (Eid, Gollwitzer & Schmitt, 2013; Urban & Mayerl, 2011). In Abhängigkeit der Ausprägung des Kriteriums lassen sich Regressionsanalysen unterteilen in lineare Regressionen für intervallskalierte Regressanden und in logistische Regressionen für Regressanden mit dichotomer bzw. kategorialer Ausprägung (Field, 2009). Weitere Modellunterschiede bestehen auf Basis der Anzahl der Prädiktoren, dem Zeitpunkt ihrer Einführung in das Regressionsmodell sowie ihrer strukturellen Beziehung zueinander (Eid et al., 2013). Eine Übersicht in Form eines Entscheidungsbaums zur Auswahl eines geeigneten Analyseverfahren in Abhängigkeit des Skalenniveaus und der Anzahl der abhängigen und unabhängigen Variablen zur Untersuchung von Zusammenhängen gibt beispielsweise Field (2009).

Das Prinzip der Regressionsanalyse lässt sich an einer einfachen linearen Regression verdeutlichen. In einer einfachen linearen Regression wird der Regressand (Y) anhand nur eines Prädiktors (X) vorhergesagt. Ihre Beziehung lässt sich mathematisch als lineare Gleichung wie folgt darstellen:

$$Y = b_0 + b_1 X_i + \varepsilon_i$$

Dabei repräsentiert der Regressionskoeffizient b_0 , auch Konstante genannt, den Schnittpunkt der Geraden mit der Y-Achse im Koordinatensystem. Der Regressionskoeffizient b_1 symbolisiert die Geradensteigung, die die Zusammenhangsstärke und Richtung zwischen dem Prädiktor und dem Kriterium definiert. Ein positiver Regressionskoeffizient b_1 stellt einen positiven Zusammenhang zwischen der unabhängigen und der abhängigen Variablen dar. Steigt der Prädiktor um eine Einheit, steigt das Kriterium um den Faktor b_1 . Ist der Regressionskoeffizient jedoch negativ, zeigt dies einen negativen Zusammenhang zwischen den beiden Faktoren. Mit jeder Einheit, die der Prädiktor ansteigt, sinkt das Kriterium um den Faktor $-b_1$. Mit ε_i wird ein Fehlerterm in die Gleichung integriert, der die nicht-erklärte Streuung im Kriterium berücksichtigt. Diese wird nicht durch Veränderungen des Prädiktors verursacht, sondern beruht beispielsweise auf weiteren, im Modell unberücksichtigten Faktoren oder Ungenauigkeiten des Messinstruments (Field, 2009; Urban & Mayerl, 2011).

In Kapitel 9 soll der subjektiv wahrgenommene Stand der beruflichen Orientierung der Jugendlichen anhand mehrerer Prädiktoren – dem Geschlecht, der Muttersprache, der Schulleistung und der besuchten berufsorientierenden Maßnahmen – geschätzt werden. Für diese Analyse wird eine hierarchische multiple Regressionsanalyse genutzt. Analog zur einfachen linearen Regression werden in einer multiplen Regression die Effektgröße und Zusammenhangsrichtung in Hinblick auf das Kriterium für alle Prädiktoren berechnet (Döring & Bortz, 2016). Für jeden Prädiktor wird also ein separater Steigungsparameter b_i benötigt. Die

mathematische Formel erweitert sich dementsprechend um die zusätzliche Anzahl an Prädiktoren:

$$Y = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + \varepsilon_i$$

Der Terminus *hierarchisch* bezieht sich auf den Einführungszeitpunkt der Variablen in das Modell. In hierarchischen Modellen werden die Prädiktoren in mehreren Schritten in das Modell integriert. Diese Vorgehensweise eignet sich besonders, um Unterschiede in der Varianzaufklärung verschiedener Modelle sichtbar zu machen (Döring & Bortz, 2016; Field, 2009). Hinsichtlich der Methodenauswahl merken Döring und Bortz (2016) an, dass Strukturgleichungsmodelle in neueren Entwicklungen den multiplen Regressionsanalysen zur Untersuchung vielschichtiger Zusammenhänge zum Teil vorgezogen werden. Dies ist insbesondere dann der Fall, wenn Zusammenhänge zwischen Variablen unterschiedlicher struktureller Ebene wie individuums- und klassenbezogenen Variablen analysiert werden sollen (Döring & Bortz, 2016). Die in Kapitel 9 dargestellten Analysen beziehen sich jedoch ausschließlich auf die Individualebene, sodass der methodische Zugang über multiple lineare Regressionsanalysen schlüssig ist.

Interaktionseffekte im Regressionsmodell

Anhand multipler Regressionsanalysen können neben den Haupteffekten der einzelnen Prädiktoren zudem ihre Interaktionseffekte, also die Effekte ihres Produkts, auf das Kriterium vorausgesagt werden (Jaccard & Turrisi, 2003). In Kapitel 9 wird die Wirkung von jeweils zwei bzw. drei Prädiktoren in direkter Abhängigkeit voneinander untersucht. Es sollen also Aussagen getroffen werden, inwiefern sich beispielsweise der (statistische) Effekt einer berufsorientierenden Maßnahme in Abhängigkeit des Geschlechts bzw. des Geschlechts und der Muttersprache auf den subjektiven Stand der beruflichen Orientierung verändert. Es handelt sich dabei um Zweifach- und Dreifachinteraktionseffekte, die nach Jaccard und Turrisi (2003) als Produkte mit jeweils einem eigenen Regressionskoeffizienten wie folgt in die Regressionsgleichung eingefügt werden:

$$\begin{aligned} Y = & (b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} \\ & + b_5X_{1i}X_{2i} + b_6X_{1i}X_{3i} + b_7X_{1i}X_{4i} + b_8X_{2i}X_{3i} + b_9X_{2i}X_{4i} + b_{10}X_{3i}X_{4i} \\ & + b_{11}X_{1i}X_{2i}X_{3i} + b_{12}X_{1i}X_{2i}X_{4i} + b_{13}X_{2i}X_{3i}X_{4i}) + \varepsilon_i \end{aligned}$$

Vorgehensweise und Voraussetzung der Daten

Nach Festlegung des zu schätzenden Kriteriums und der zu verwendenden Prädiktoren anhand eines theoretischen Modells bzw. vorangegangener empirischer Studien, wird die Reihenfolge festgelegt, in der die Prädiktoren in das Modell eingeführt werden (Field, 2009). Dabei werden nach Field (2009) im Idealfall Prädiktoren, deren Effekte durch vorangegangene Studien bereits erklärt wurden, vor neu zu betrachtenden Prädiktoren in das Modell aufgenommen. In Kapitel 9 werden daher in einem ersten Schritt die beiden soziodemografischen Merkmale Geschlecht und Muttersprache sowie die Schulleistung eingeführt. Im zweiten Schritt wird das Modell durch den Prädiktor *berufsorientierende Maßnahme* erweitert. Im dritten Schritt werden alle anhand der vier Prädiktoren möglichen Zweifachinteraktionen ergänzt, um schließlich im vierten Schritt durch den Zusatz der Dreifachinteraktionen das finale Modell zu erhalten.

Die Voraussetzungen zur Durchführung und Interpretation einer multiplen Regressionsanalyse werden vorab sowie als Teil einer ersten Regressionsrechnung überprüft. Als Erstes werden verzerrend wirkende Ausreißer in den Daten mittels eines Streudiagramms der unabhängigen und abhängigen Variablen visuell identifiziert (Field, 2009). Darüber hinaus werden die Daten anhand der Residualgrößen, also der Differenz zwischen den erwarteten und den beobachteten Ergebnissen, auf Verzerrungen hin untersucht (Field, 2009). Verzerrungen im Regressionsmodell, die durch Fälle mit zu starkem Einfluss verursacht werden, können anhand des Cook'schen Distanzmaßes, der Hebelwerte sowie der multivariaten Mahalanobis-Distanz ermittelt werden (Field, 2009). Field (2009) empfiehlt schließlich noch den Einfluss eines Falls auf die Varianz der Regressionskoeffizienten mittels seiner Covarianz Ratio (CVR) zu überprüfen. Für identifizierte Ausreißer muss dann – auch auf Basis theoretischer Überlegungen – entschieden werden, ob sie von der Analyse ausgeschlossen oder beibehalten werden sollen.

Vor der Interpretation des Regressionsmodells müssen weitere Voraussetzungen sichergestellt werden. Eine zentrale Annahme der linearen Regression ist die lineare Beziehung zwischen den Prädiktoren und dem Kriterium (Field, 2009). Diese Voraussetzung wird anhand des Residuen-Streuungsdiagramms validiert. Lässt sich ein Muster im Streudiagramm erkennen, deutet dies auf Heteroskedastizität (siehe nächster Abschnitt) und eine Nichtlinearität des Zusammenhangs zwischen Prädiktoren und Kriterium hin (Field, 2009). Anhand des Durbin-Watson Tests wird die Unabhängigkeit der Residuen – oder in anderen Worten der Ausschluss einer Autokorrelation – als weitere Voraussetzung von linearen Regressionsanalysen überprüft (Field, 2009). Autokorrelationen können aufgrund eines nicht-linearen Zusammenhangs vorliegen oder darin begründet sein, dass erklärungsrelevante Prädiktoren im Modell fehlen (Field, 2009). Als Konsequenz wären die Konfidenzintervalle

und die Signifikanztests ungültig. Für die Gültigkeit der Konfidenzintervalle und der Signifikanztests ist zudem die Homoskedastizität der Residuen von Bedeutung. Auch wenn nach Simulationsstudien von Lumley, Diehr, Emerson und Chen (2002) Regressionen robust auf diese Voraussetzungsverletzung reagieren, sollte das Vorliegen von Heteroskedastizität, also einer sich verändernden Varianz der Residuen der Prädiktoren, anhand eines Residuen-Streuungsdiagramms visuell überprüft und ausgeschlossen werden (Field, 2009). Als dritte bedingende Verteilungseigenschaft der Residuen muss die Normalverteilung der Residuen visuell anhand ihres Histogramms sowie eines Quantil-Quantil-Diagramms (Q-Q-Plot) bestätigt werden. Auch hier reagiert die Regression nach Backhaus, Erichson, Plinke und Weiber (2016) ab einer Stichprobengröße von $N > 40$ robust auf eine Verletzung der Voraussetzung. Zuletzt darf keine Multikollinearität der Residuen vorliegen, also keine direkte lineare Abhängigkeit der Prädiktoren untereinander. Korrelieren die Prädiktoren zu stark untereinander, lässt sich im Regressionsmodell ihr jeweiliger Anteil an der erklärten Varianz nicht klar trennbar messen. Zur zusätzlichen Überprüfung eignen sich zudem die Toleranzwerte sowie die Varianz-Inflations-Faktoren (Field, 2009; Urban & Mayerl, 2011).

Modellgüte und Effektstärken

Zur Bestimmung der Modellgüte, also wie gut ein Modell die beobachteten Daten voraussagt, werden für die hierarchische multiple Regression der Determinationskoeffizient R^2 , die Änderungen in R^2 sowie die F-Statistik herangezogen. R^2 drückt den Anteil der Varianz im Kriterium aus, der durch die Prädiktoren erklärt wird (Field, 2009). Um die vier Modellschritte miteinander vergleichen und daraus abzuleiten zu können, inwiefern das Hinzufügen weiterer Variablen die Varianzaufklärung verbessert, werden die Änderungen in R^2 der Modelle betrachtet. Ein signifikanter p -Wert der F-Statistik indiziert eine Allgemeingültigkeit des Regressionsmodell für die Gesamtpopulation auch über die Stichprobe hinaus (Field, 2009). Neben der Validität des Gesamtmodells ist der Beitrag der einzelnen Prädiktoren von hohem Interesse. Der unstandardisierte Regressionskoeffizient b gibt sowohl den Anteil des Prädiktors am Gesamtmodell als auch die Beziehung zwischen dem Prädiktor und dem Kriterium an. Ein positiver Wert für b weist auf eine positive Beziehung der beiden Variablen hin: steigt der Prädiktor um eine Einheit, steigt auch der Wert der abhängigen Variablen um b Einheiten. Bei einem negativen b -Wert ist die Beziehung gegenläufig, sodass die Werte des Kriteriums mit steigendem b sinken (Field, 2009). Ist der t -Wert des jeweiligen Prädiktors signifikant, kann sein Beitrag zum Gesamtmodell als bedeutsam eingestuft werden. Um die Effekte der einzelnen Prädiktoren miteinander vergleichen zu können, werden ihre standardisierten Regressionskoeffizienten, Beta (β) miteinander verglichen. Sie stellen den Regressionskoeffizienten

in Standardabweichungen dar: Am Wert von Beta wird demnach abgelesen, um wie viele Standardabweichungen sich das Kriterium verändert, wenn der Prädiktor um eine Standardabweichung ansteigt (Field, 2009).

Statistische Verfahren zur Normwertbestimmung in Kapitel 10

Sollen die Testergebnisse einer einzelnen Person in Relation zu den Testergebnissen ihrer Peergruppe interpretiert werden, eignet sich dafür die Nutzung normorientierter Tests. Im Folgenden soll das Vorgehen der empirischen Studie in Kapitel 10 zur Testung eines zentralen Berufswahlkompetenz-Normwerts erläutert werden.

Ziel und Prinzip von Normwertbestimmungen

Normorientierte Testauswertungen zielen darauf ab, das Ergebnis eines einzelnen Individuums im Kontext seiner Peergruppe interpretieren zu können (Bühner, 2011). Für diese Ergebniseinordnung bedarf es zuvor einer normierten Eichstichprobe, auch Normierungsstichprobe genannt, die die Merkmale der Grundgesamtheit der Testzielgruppe repräsentativ abbildet (Döring & Bortz, 2016). Ziel dieser Normierung ist die Ableitung von Grenzwerten, auch *Cut-Off*-Werte genannt, die wiederum die Eingruppierung von Testergebnissen und folgend ihre Interpretation ermöglichen. Soll ein Test beispielsweise die Lesekompetenz von Lernenden in der Grundschule erfassen, bildet die entsprechende Eichstichprobe die Gesamtheit der Grundschüler*innen hinsichtlich ihres Geschlechts, Alters, ihrer Herkunft, ihres sozioökonomischen Status und anderer Merkmale ab.

Ein weiteres avisiertes Ziel der Nutzung von Normwerten kann zudem die Komplexitätsreduzierung eines Gesamtkonstrukts, das sich aus verschiedenen Unterkategorien zusammensetzt, sein. Denn ein Normwert sollte – auch im Sinne der Testökonomie, also dem angemessenen Verhältnis zwischen dem Erkenntnisgewinn und den aufzuwendenden Ressourcen (Döring & Bortz, 2016) – diese Unterkategorien bzw. Facetten abbilden (Scharfetter, 1996). Die Komplexitätsreduzierung ist insbesondere nützlich, wenn eine Vielzahl an Personen zur diagnostischen Einschätzung an einem standardisierten Test teilnehmen und ihre Ergebnisse ausgewertet werden sollen. Das wahrscheinlich bekannteste Beispiel eines Normwerts ist die IQ-Norm, über die die Intelligenz eines Individuums zeitökonomisch erfasst werden soll (Bühner, 2011). Gleichzeitig ist die IQ-Norm ein Beispiel für einen übergreifenden Wert, der die verschiedenen Facetten eines Gesamtkonstrukts integriert und repräsentiert. In Abhängigkeit des Intelligenztests basiert der ermittelte IQ auf Ergebnissen aus unterschiedlichen Teilbereichen von Intelligenz wie im Falle des WAIS-IV (*Wechsler Adult Intelligence Scale IV*) auf dem Sprachverständnis, dem logischen Denken, dem Arbeitsgedächtnis und der Verarbeitungsgeschwindigkeit des Testsubjekts (Petermann, 2012).

In der in Kapitel 10 vorgestellten empirischen Studie wird der Bedarf von Schulen hinsichtlich eines zeit- und ressourceneffizienten Verfahrens zur Feststellung individueller Bedarfe in der Berufswahlkompetenzentwicklung adressiert (Ohle- mann et al., 2016). Die Studie untersucht, inwiefern ein zentraler Normwert der Berufswahlkompetenz gebildet werden kann und inwiefern dieser den Entwicklungsstand auf den zwölf Facetten von Berufswahlkompetenz abbildet. Dabei ist wichtig anzumerken, dass es sich bei dem verwendeten Verfahren zur Diagnostik von Berufswahlkompetenz nicht um einen kognitiven Leistungstest handelt wie beispielsweise bei der Erfassung des IQs, sondern um ein Selbsteinschätzungsverfahren, bei dem die Teilnehmenden ihr berufsbezogenes Wissen, ihre Motivation und Handlungskompetenz selbst bewerten (Kaak et al., 2013).

Vorgehensweise und Voraussetzung der Daten

Zur Normierung eines Tests werden aus den oben beschriebenen Gründen in einem ersten Schritt die Testergebnisse einer repräsentativen Stichprobe untersucht. Unter der Voraussetzung einer Normalverteilung können die Skalenrohwerte der Eichstichprobe in z -Werte überführt werden. Bei der z -Transformation wird der Skalenmittelwert jeweils vom Skalenrohwert subtrahiert und durch die Standardabweichung geteilt. Es entsteht eine gleichmäßige, glockenförmige Verteilung mit einer Standardabweichung von 1 und einem Mittelwert von 0 (Döring & Bortz, 2016). In einem zweiten Schritt können die z -Werte in die gewünschte Normskala, beispielsweise in die C-Skala, überführt werden. Abbildung 8.1 stellt eine Auswahl gebräuchlicher Normskalen dar. Darin wird u. a. ersichtlich, dass das Spektrum der C-Skala zwischen einem Minimum von -1 und einem Maximum von 11 liegt.

Die in Kapitel 10 betrachteten zwölf Facetten von Berufswahlkompetenz (Driesel-Lange et al., 2020) folgen jedoch mehrheitlich keiner Normalverteilung. Ebenso ist der aus ihnen gebildete Mittelwert nicht normalverteilt, sodass eine Transformation in eine Normskala einer gesonderten Bearbeitung der Skalenrohwerte bedarf (Bühner, 2011). Für die Normierung nicht-normalverteilter Werte eignen sich Staninewerte, da „sie die schiefe Verteilung unter Umständen normalisieren können“ (Bühner, 2011, S. 265). Das Wort *Stanine* setzt sich aus den englischen Wörtern *standard* und *nine* zusammen und stellt eine auf neun Werte ($Min = 1$; $Max = 9$) reduzierte C-Skala mit einem Mittelwert von fünf dar (Bühner, 2011, siehe auch Abbildung 8.1). Staninewerte werden, wie Bühner (2011) spezifiziert, häufig zur Normierung von Persönlichkeitstests genutzt.

Döring und Bortz (2016) beschreiben die Umwandlung von Skalenrohwerten in Staninewerte wie folgt: Zuerst werden die Skalenrohwerte vom niedrigsten bis zum höchsten Wert aufsteigend geordnet. Die Platzierung eines Wertes wird als Prozen- trangwert bezeichnet (Döring & Bortz, 2016). Nun werden die Werte entsprechend

Abbildung 8.1 Skizzierung ausgewählter Normen in Anlehnung an Bühner (2011).
Anmerkung. HAWIE = Hamburg-Wechsler-Intelligenztest für Erwachsene

ihres Prozentranges einer der neun Staninegruppen zugeordnet. Tabelle 8.2 zeigt die Prozentbereiche und ihre jeweilige Zuordnung zu einer Staninegruppe. Die untersten 4 % der Skalenrohwerte werden der ersten Staninegruppe zugeordnet, die folgenden 7 % bzw. 12 % der zweiten und dritten Staninegruppe.

Tabelle 8.2 Umrechnung von Prozenträngen in Staninewerte in Anlehnung an Döring und Bortz (2016)

Testwert-Rang	Prozentrang	Staninegruppe
Niedrigste 4 %	0 – 4	1
7 %	> 4 – 11	2
12 %	> 11 – 23	3
17 %	> 23 – 40	4
20 %	> 40 – 60	5
17 %	> 60 – 77	6
12 %	> 77 – 89	7
7 %	> 89 – 96	8
Höchste 4 %	> 96 – 100	9

Wie aus der Tabelle 8.2 abzulesen ist, werden die folgenden 17 % der Staninegruppe vier zugeordnet. Die fünfte Gruppe, in der die mittleren 20 % der Rohwerte liegen, bildet den mittleren Bereich der Stanineskala. Die Staninegruppen sechs bis neun spiegeln die prozentualen Gruppengrößen der ersten vier Staninegruppen im oberen Spektrum (Döring & Bortz, 2016).

Nach der Einteilung der Eichstichprobe in die Staninegruppen können für jede Staninegruppe die Mittelwerte des Normwerts und seiner Facetten sowie die entsprechenden 95 % Konfidenzintervalle des Standardfehlers des Mittelwerts (*KI*) berechnet werden. Das 95 % Konfidenzintervalle des Standardfehlers des Mittelwerts wird folgendermaßen berechnet:

$$KI = M \pm 1.96 \times SE$$

Die Werte der 95 % Konfidenzintervalle können als *Cut-Off*-Werte für die entsprechenden Gruppen herangezogen werden. Alternativ kann zuvor eine Aggregation der Staninegruppen in weniger Gruppen zur weiteren Komplexitätsreduzierung, beispielsweise für die Anwendung in der schulischen Praxis, vorgenommen werden. In Kapitel 10 wurden jeweils drei benachbarte Staninegruppen zusammengefasst, nämlich jeweils die Staninegruppen eins bis drei, die Staninegruppen vier bis sechs und die Staninegruppen sieben bis neun. Bedeutsam für eine Anwendung in der Praxis ist dabei auch die Trennschärfe der Gruppen. Über die 95 % Konfidenzintervalle kann diese hinsichtlich des Normwerts, aber auch in Bezug auf einzelne Unterkategorien überprüft werden (Schurtz & Artelt, 2014). Überschneiden sich die 95 % Konfidenzintervalle der Gruppen nicht, handelt es sich um eine rein differenzierbare Facette. Das Gegenteil, also eine undifferenzierte Facette, liegt vor, wenn sich die 95 % Konfidenzintervalle zweier oder mehrerer Gruppen überlagern (Schurtz & Artelt, 2014). Eine annähernde Differenzierung des Normwerts besteht, wenn die Gruppen offenbar klare Unterschiede in ihren Profilen zeigen, jedoch Überlappungen der Gruppen auf einer oder mehreren Facetten vorliegen (Schurtz & Artelt, 2014).

Klassifizierungsverfahren in Kapitel 11: Latente Profilanalysen

Zur Identifizierung von Datenmustern und der Erstellung von Klassifizierungen lassen sich verschiedene Klassifizierungsverfahren heranziehen. Im Folgenden soll die in Kapitel 11 verwendete latente Profilanalyse (*latent profile analysis*, LPA) im Detail vorgestellt und zu anderen Klassifizierungsverfahren abgegrenzt werden.

Ziel und Prinzip der latenten Profilanalyse

Die Theorie latenter Klassen (Dayton & Macready, 1976; Graham, Collins, Wugalter, Chung & Hansen, 1991; Lazarsfeld & Henry, 1968) beruht als Messtheorie auf der Annahme, dass die Population auf Basis nicht direkt messbarer Muster in verschiedene, sich gegenseitig ausschließende latente Gruppen unterteilt werden kann (Graham et al., 1991). Die Methode der latenten Klassenanalyse geht auf Lazarsfeld und Henry (1968) zurück und basiert auf eben dieser Annahme.

Ziel der latenten Klassenanalyse (*latent class analysis*, LCA) ist es demnach, latente Muster zu identifizieren und Daten mit ähnlichen Datenmustern zu homogenen Subgruppen oder Klassen zusammenzuführen (Nylund, 2007). Döring und Bortz (2016) definieren latente Merkmale als Eigenschaften, die nicht direkt zu beobachten und dadurch im Gegensatz zu manifesten Variablen wie dem Alter einer Person oder der Schulklasse folglich nicht direkt messbar sind. Die latente Profilanalyse (*latent profile analysis*, LPA) bezeichnet eine Anwendungsform der latenten Klassenanalysen für metrisch skalierte Indikatoren (Vermunt, 2004).

Die Untersuchung versteckter Muster kann beispielsweise zur empirischen Bestätigung von Theorien, der Erstellung von Typologien oder von diagnostischen Instrumenten genutzt werden (Geiser, 2010; Vermunt, 2004). Auch Faktoranalysen werden, ähnlich den latenten Klassen- und Profilanalysen, zur Theorieüberprüfung und Identifikation latenter Strukturen verwendet. Allerdings werden anhand von Faktorenanalysen latente kontinuierliche Variablen untersucht (Nylund, 2007) und Items zu homogenen Subgruppen, den Faktoren, zusammengefasst (Döring & Bortz, 2016). Es handelt sich bei der Faktoranalyse insofern um eine Itemklassifizierung. Im Gegensatz dazu sind latente Klassen- und Profilanalysen individuumszentrierte Verfahren, bei denen einzelne Datenfälle zu homogenen Subgruppen gebündelt werden. Die aus latenten Klassen- und Profilanalysen entstehenden Ergebnisvariable ist kategorial (Nylund, 2007).

Auch traditionelle Clusteranalysen, wie das *K-means*-Verfahren, haben die Bündelung individueller Fälle, denen ein ähnliches Datenmuster zu Grunde liegt, zum Ziel (DiStefano & Kamphaus, 2006). Der wesentliche Unterschied zwischen traditionellen Clusterverfahren und latenten Klassen- und Profilanalysen besteht in dem modellbasierten Ansatz der LCA/LPA (Magidson & Vermunt, 2002). Latente Analysen erstellen auf Basis von Wahrscheinlichkeitsverteilungen ein für die Population, aus der die Stichprobe stammt, geltendes statistisches Modell (Magidson & Vermunt, 2002). Zentral in diesem Modell sind die *item parameters*, die die item-bezogenen Wahrscheinlichkeiten für jede Profilgruppe (Mittelwerte und Varianzen) angeben, sowie die *class probability parameters*, die wiederum die Prävalenz der jeweiligen Profilgruppe bezeichnen (Nylund, Asparouhov & Muthén, 2007). Die mathematische Formel zur Bestimmung latenter Profile lautet nach Nylund et al. (2007) wie folgt:

$$f(y_i) = \sum_{k=1}^K P(c = k) f(y_i | c = k)$$

y_i stellt dabei den Vektor der Antworten einer Person i dar, P bezeichnet die Wahrscheinlichkeit und c steht für die latente Variable mit k Klassen (Nylund et al., 2007). Das Verfahren latenter Klassen- und Profilanalysen ist nach Magidson und Vermunt (2002) den traditionellen Clustermethoden vorzuziehen, da sie laut ihrer Simulationsstudien in stabileren latenten Subgruppen resultieren.

Vorgehensweise und Voraussetzung der Daten

Zuerst bedarf es einer theoretischen Zuordnung der zu untersuchenden latenten Variable, also einer Definition, auf welchem theoretischen Konstrukt diese fußt bzw. welches theoretische Konstrukt die latente Variable abbilden soll (Döring & Bortz, 2016). Im Anschluss an diese Konzeptspezifikation werden die manifesten Variablen – auch Indikatoren genannt –, über die die latente Variable gemessen werden soll, entsprechend bestimmt (Vermunt, 2004). Für die manifesten Variablen wird eine Normalverteilung angenommen (Geiser, 2010).

In Kapitel 11 sollen latente Berufswahlkompetenzprofile ermittelt werden. Diese fundieren auf dem Berufswahlkompetenzmodell von Driesel-Lange et al. (2010). Als Indikatoren werden die zwölf Berufswahlkompetenzfacetten herangezogen. Nach Definition der latenten Variablen und Bestimmung der Indikatoren werden mehrere LPA-Modelle mit steigender Profilanzahl beginnend mit einer Profilgruppe berechnet.

Die Full Information Maximum Likelihood-Methode (FIML) ermöglicht alle verfügbaren Daten einzubeziehen, auch solche mit zufallsbedingt punktuell fehlenden Werten (Geiser, 2010; Graham, 2012). Für jedes LPA-Modell wird die Zuordnung der einzelnen Fälle zu den latenten Profilen als Ergebnisvariable ausgegeben und für spätere Analysen gespeichert. Ein Ziel der LPA im Hinblick auf die Modellgüte besteht darin, das Modell mit dem größten Log-Likelihood-Wert zu wählen. Zur Vermeidung eines lokalen Likelihood-Maximums sollte eine ausreichend große Anzahl an zufällig generierten Startwertesets mit entsprechend zahlreichen Iterationen gewählt werden (Geiser, 2010). McLachlan und Peel (2000) empfehlen daher die Anzahl der Startwertesets für eine erste Analyse auf 500 Startwertesets mit 50 Iterationen zu fixieren, um lokale Wahrscheinlichkeitsmaxima zu vermeiden. Eine zweite Analyserunde zur Bestätigung der gefundenen Modellparameter sollte mit 1000 Startwertesets mit 100 Iterationen durchgeführt werden (vgl. Geiser, 2010). Die LPA-Modelle werden dann auf ihre Modellgüte und Profilschärfe hin verglichen, um das finale LPA-Modell und damit die Anzahl der Profile zu bestimmen (Nylund et al., 2007).

Modellgüte und Auswahl der Profile

Zur Bestimmung des Modells, das die latenten Strukturen der Daten am besten abbildet, und der damit einhergehenden Bestimmung der Anzahl der Profile werden zunächst die Modellfit-Indizes der einzelnen Modelle betrachtet und miteinander verglichen. Bei LCAs mit binären manifesten Variablen wird der absolute Modellfit mit Hilfe des Loglikelihood-Ratio-Tests sowie des Pearson- χ^2 -Tests bestimmt (Geiser, 2010). Der absolute Modellfit sagt aus, inwiefern das jeweilige Modell die erfassten latenten Strukturen darstellt (Geiser, 2010). Aufgrund der metrischen Skalierung der manifesten Variablen stehen die genannten Verfahren für latente Profilanalysen jedoch nicht zur Verfügung. Aber es besteht die Möglichkeit den relativen Modellfit zu bewerten. Dazu können mit Hilfe der informationstheoretischen Maße die verschiedenen Modelle miteinander verglichen werden.

In Abhängigkeit der Stichprobengröße und des Modells empfehlen Autor*innen verschiedener Simulationsstudien (u. a. Nylund et al., 2007; Tein, Coxé & Cham, 2013) den Einsatz des Akaike Information Criteria (AIC), des Bayesian Information Criteria (BIC) oder des sample-size adjusted BIC (SaBIC). Allen drei Indizes ist gemeinsam, dass sie Modellgüte gegen Modellsparsamkeit abwägen (Geiser, 2010). Bei ausschließlicher Betrachtung der informationstheoretischen Maße sollte das Modell mit dem jeweils kleinsten Wert für den AIC, BIC bzw. SaBIC ausgewählt werden (Geiser, 2010). Tein et al. (2013) empfehlen auf Basis ihrer Simulationsstudien für mittlere bis große Stichproben ($N > 250$) die Nutzung des SaBICs zur Bestimmung der Profilanzahl und raten von der Interpretation des AICs aufgrund seiner geringen Reliabilität ab.

Neben diesen deskriptiven Fit-Indizes können zur Auswahl der Profilanzahl verschiedene statistische Tests herangezogen werden: Der Bootstrapp-Likelihood-Ratio-Differenztest (BLRT) vergleicht jeweils ein Modell mit k Profilen mit dem nächstkleineren Modell, also mit $k - 1$ Klassen. Ist der Wert des BLRTs signifikant, deutet dies auf eine bessere Datenpassung des Modells mit k Klassen hin (Geiser, 2010). Der Vuong-Lo-Mendell-Rubin-Test (VLMRT) basiert auf einem ähnlichen Prinzip. Auch hier deutet ein signifikanter Wert der VLMRTs darauf hin, dass das geschätzte Modell die Daten besser abbildet als das nächst kleinere Modell (Geiser, 2010).

Die geschätzte mittlere Klassenzuordnungswahrscheinlichkeit gibt an, mit welcher Wahrscheinlichkeit ein Fall auf Basis seines Datenmusters dem passenden Profil korrekt zugeordnet wird (Geiser, 2010). Insbesondere für den Einsatz von Profilen in der Diagnostik stellt die Klassenzuordnungswahrscheinlichkeit ein wichtiges Qualitätsmaß dar. Würden die in Kapitel 11 identifizierten Berufswahlkompetenzprofile anschließend zur Diagnostik weiterer Orientierungs- und Beratungsbedarfe Jugendlicher verwendet werden, wäre eine korrekte Profilzuordnung von besonders

hoher Bedeutung. Rost (2006) empfiehlt als Daumenregel bei einem maximalen Wert von 1, der einer 100 Prozent richtigen Zuordnung entsprechen würde, einen Wert von größer 0.8 für alle Profile. Entsprechend liegt die Wahrscheinlichkeit der korrekten Profilzuordnung bei 80 Prozent (Geiser, 2010). Mit steigender Profilzahl verbessern sich auch die Werte der Fit-Indizes, da eine große Anzahl an Profilen komplexe Datenstrukturen besser als Modelle mit wenigen Profilgruppen im Detail abbilden können. Gleichzeitig sinkt mit der Anzahl der Profile auch die Größe der einzelnen Klassen und damit ihre Anwendbarkeit in der Praxis. Insofern wird die Klassengröße, geschätzt anhand des Klassengrößenparameters $\hat{\pi}_g$, als Kriterium eines Empirie-Praxis-Transfers in der Modellauswahl berücksichtigt, um Profile mit nur wenigen Fallzahlen zu vermeiden (Geiser, 2010).

Für die Auswahl des finalen Modells soll neben der Auswertung der beschriebenen statistischen Werte auch die Interpretierbarkeit der Profile berücksichtigt werden (Nylund-Gibson, Grimm, Quirk & Furlong, 2014; Nylund et al., 2007). Um eine theoriebezogen sinnvolle Interpretation sicherzustellen (Wang & Wang, 2012), wurden für den Modellvergleich in Kapitel 11 daher auch theoretische Annahmen des Berufswahlkompetenzmodells von Driesel-Lange et al. (2010) berücksichtigt.

Die Entropie als Gütemaß der Klassifikation eignet sich nach Simulationen von Celeux und Soromenho (1996) nicht zur Modellauswahl, denn ihre Werte basieren auf der Hypothese, dass das korrekte Modell ausgewählt wurde. Dennoch kann nach Auswahl eines Modells die Entropie zur Interpretation der Zuverlässigkeit der Klassifikation herangezogen werden (Nylund-Gibson et al., 2014). Die Entropie nimmt Werte zwischen 0 und 1 an. Je näher der Entropiewert an 1 liegt, desto höher ist die Wahrscheinlichkeit, dass die einzelnen Fälle der korrekten Profilgruppe zugeordnet werden (Geiser, 2010). Zur weiteren Interpretation der latenten Ergebnisvariable, also der Profile, werden zusätzliche Variablen etwa soziodemografische Faktoren oder persönliche Einstellungen herangezogen (Geiser, 2010). In Kapitel 11 werden die Berufswahlkompetenzprofile in Hinblick auf den Einfluss des Geschlechts, des Alters und der (In-)Existenz eines konkreten Berufswunschs der Jugendlichen auf die jeweilige Profilzuordnung hin untersucht. Hierfür wurde eine multinomiale logistische Regression verwendet, deren Verfahren im Folgenden erläutert wird.

Zusammenhangsanalysen in Kapitel 11: multinomiale logistische Regression

Wie bereits zu Beginn des Kapitels erläutert, können Zusammenhänge zwischen mehreren Variablen anhand von Regressionsanalysen untersucht werden (Döring & Bortz, 2016). Im Folgenden soll nun das Prinzip der logistischen Regression dargestellt werden, das in Kapitel 11 Anwendung findet, um anschließend die Voraussetzungen und Vorgehensweise zu erläutern.

Ziel und Prinzip der logistischen Regression

Im Unterschied zu linearen Regressionsmodellen, die nur auf metrisch skalierte Kriterien angewendet werden können, können mittels logistischer Regressionen Einflüsse von Prädiktoren auf ein nominalskaliertes Kriterium geschätzt werden (Field, 2009). Ist das Kriterium dichotom ausgeprägt, wird der Zusammenhangseffekt anhand einer binär-logistischen Regression geschätzt. In Kapitel 11 soll jedoch der Effekt des Geschlechts, des Alters und des Berufswunschs auf die Zugehörigkeit zu den vier identifizierten Berufswahlkompetenzprofilen, also einem Kriterium mit vier Ausprägungen, untersucht werden. Zur Prädiktion eines solchen polytom-nominalskalierten Kriteriums wird ein multinomiales logistisches Regressionsmodell herangezogen (Eid et al., 2013). Multinomiale logistische Regressionen gehören wie multiple lineare Regressionen zu den multivariaten Verfahren, die im Gegensatz zu bivariaten Modellen die Bedeutung von Merkmalsbeziehungen aufnehmen (Döring & Bortz, 2016).

Im logistischen Regressionsmodell wird nicht von einem linearen Zusammenhang zwischen den Prädiktoren und dem Kriterium ausgegangen, sondern anhand einer logistischen Funktion ein s-förmiger Zusammenhang modelliert (Backhaus et al., 2016). Im Unterschied zur linearen Regression wird dabei nicht der Wert des Kriteriums, sondern die Wahrscheinlichkeit geschätzt, dass eine der jeweiligen Ausprägungen des Kriteriums eintritt (Field, 2009). Das Ergebnis wird als Zufall behandelt (Backhaus et al., 2016). Die mathematische Gleichung zur Wahrscheinlichkeitsrechnung sieht, wie Field (2009) zeigt, folgendermaßen aus:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_i)}}$$

Dabei stellt P die Wahrscheinlichkeit des Eintretens einer Ausprägung des Kriteriums (Y) dar. P kann nur Werte zwischen 0 (die Ausprägung tritt nicht ein) und 1 (die Ausprägung tritt ein) annehmen. Wie bei der linearen Regression stellen b_0 und b_1 jeweils die Konstante beziehungsweise den Regressionskoeffizienten der unabhängigen Variablen dar. In Kapitel 11 wird jeweils die Wahrscheinlichkeit geschätzt, mit der eine Person aufgrund der Existenz bzw. des Fehlens eines konkreten Berufswunschs zu einem Berufswahlkompetenzprofil zugeordnet wird. Um die Modellpassung zu schätzen, wird – analog zur linearen Regression – auf den Vergleich der beobachteten und vorhergesagten Werte zurückgegriffen. Jedoch erfolgt dies aufgrund des fehlenden linearen Zusammenhangs mittels einer Log-Likelihood-Funktion:

$$\text{Log-Likelihood} = \sum_{i=1}^N [Y_i \ln(P(Y_i)) + (1 - Y_i) \ln(1 - P(Y_i))]$$

Diese misst die Varianz im Modell, die nicht durch den Prädiktor erklärt wird (Field, 2009).

Vorgehensweise und Voraussetzung der Daten

Wie auch bei der linearen Regression soll bei der logistischen Regression ein möglichst sparsames Modell gewählt werden, das mit wenigen, aber einflussreichen Prädiktoren auskommt (Backhaus et al., 2016). Für die Wahl des sparsamsten Modells werden zu Beginn verschiedene Modelle untereinander sowie mit dem Grundmodell verglichen, in dem nur die Konstante enthalten ist (Field, 2009). Die Verteilungseigenschaften der Residuen (Inexistenz von Ausreißern und überproportional einflussnehmenden Fällen, Abwesenheit von Multikollinearität) werden wie bei der linearen Regression überprüft (siehe erster Abschnitt dieses Kapitels). Statt der linearen Beziehung des Prädiktors und des Kriteriums in der linearen Regression, stellt für die logistische Regression die Linearität des Logits eine zu testende Voraussetzung dar (Field, 2009).

Modellgüte und Effektstärke

Die globale Anpassungsgüte des Gesamtmodells lässt sich für eine logistische Regression mittels Loglikelihood-Ratio (LR) einschätzen. Nach Field (2009) weist ein niedriger LR-Wert auf eine gute Passung hin, während ein hoher Wert grundsätzlich Anlass zur Sorge bezüglich der Passung birgt. Der LR-Test dient darüber hinaus dem Vergleich der Anpassungsgüte verschiedener Modelle. Dafür wird der LR-Wert eines Modells mit dem LR-Wert des Nullmodells verglichen: Ist der Wert des Gesamtmodells kleiner als der des Nullmodells lässt sich annehmen, dass das Gesamtmodell das Kriterium besser voraussagt als das Modell mit der Konstanten (Field, 2009).

Weitere globale Gütemaße zur Einordnung des Gesamtmodells stellen die in Anlehnung an den Determinationskoeffizient R^2 benannten Pseudo- R^2 wie Cox & Snells R^2_{CS} oder Nagelkerkes R^2_N dar (Backhaus et al., 2016). Ähnlich dem LR-Test können anhand der Pseudo- R^2 verschiedene Modelle miteinander verglichen werden. Die Pseudo- R^2 vergleichen das Wahrscheinlichkeitsverhältnis zwischen einem Nullmodell und dem Gesamtmodell (Backhaus et al., 2016). Damit kann eine Aussage darüber getroffen werden, um wie viel die Schätzung der Wahrscheinlichkeit des Eintreffens einer Kriteriumsausprägung durch das Hinzufügen des Prädiktors verbessert werden kann (Mayerl & Urban, 2010). Angewandt auf die Studie in Kapitel 11 lässt sich anhand des Pseudo- R^2 eine Aussage treffen, um wieviel Prozent

der Prädiktor Berufswunsch die Schätzung der Wahrscheinlichkeit im Vergleich zum Nullmodell verbessert, dass eine Person einem bestimmten Berufswahlkompetenzprofil zugeordnet wird. Je näher der Wert für R^2 an 1 liegt, desto besser sagt das Modell die Wahrscheinlichkeit des Kriteriums voraus (Backhaus et al., 2016). Für multinomiale logistische Regressionen ist darüber hinaus die Pearson-Chi-Quadrat-Statistik von Relevanz: Kleine, nicht-signifikante Werte weisen auf eine hohe Anpassungsgüte hin (Field, 2009).

Bei der Interpretation der Regressionskoeffizienten besteht, wie Backhaus et al. (2016) erklären, ein wesentlicher Unterschied zwischen der logistischen Regression und der linearen Regression. Denn in Ersterer kann nicht von einem linearen und damit konstanten Zusammenhang zwischen der Veränderung des Prädiktors und des Kriteriums ausgegangen werden. Anhand des Regressionskoeffizienten b lässt sich die Steigung und der Verlauf des s-förmigen Zusammenhangs des Kriteriums mit der Veränderung des Prädiktors erkennen (Backhaus et al., 2016). Um den Einfluss des jeweiligen Prädiktors auf die Veränderung des Kriteriums jedoch zu interpretieren, wird seine Odds-Ratio, auch Chancenverhältnis genannt, betrachtet. Bei der Odds-Ratio handelt es sich um das Verhältnis von der Wahrscheinlichkeit, dass eine Kriteriumsausprägung eintritt, zur Wahrscheinlichkeit, dass sie nicht eintritt (Backhaus et al., 2016). Ist die Odds-Ratio eines Prädiktors größer als 1, so steigt mit einem ansteigenden Wert des Prädiktors auch die Wahrscheinlichkeit, dass die jeweilige Kriteriumsausprägung eintritt. Im Gegenzug sinkt diese Wahrscheinlichkeit mit dem Ansteigen des Prädiktorwerts, wenn seine Odds-Ratio kleiner als 1 ist (Field, 2009). Inwiefern ein Prädiktor Einfluss auf die Wahrscheinlichkeit des Eintreffens einer Kriteriumsausprägung hat, gibt in der logistischen Regression der Wald-Test an, der dem t -Test in der linearen Regression ähnelt: Weist dieser einen p -Wert von kleiner oder gleich .05 auf, kann der Einfluss des Prädiktors als signifikant angesehen werden (Backhaus et al., 2016).

Latente Transitionsanalysen zur Untersuchung längsschnittlicher Bewegungsmuster in Kapitel 12

Sind latente Muster längsschnittlicher Daten von Interesse, können diese mithilfe latenter Transitionsanalysen (*latent transition analysis*, LTA) untersucht werden. Im Folgenden soll die in Kapitel 12 verwendete latente Transitionsanalyse im Detail erläutert und zu anderen längsschnittlichen Analyseverfahren abgegrenzt werden.

Ziel und Prinzip der latenten Transitionsanalyse

Methodisch ist die latente Transitionsanalyse als latentes Markov-Modell einzuordnen, also ein Modell, das die Übergänge zwischen verschiedenen Stadien im zeitlichen Verlauf beschreibt (Nylund, 2007). Ähnlich wie latente Klassen- und

Profilanalysen, die sich für querschnittliche Daten eignen (siehe vorangegangener Abschnitt zu LPA), basiert die LTA auf der Annahme, dass unsichtbare, nicht direkt messbare Strukturen vorliegen, die mittels manifester Variablen sichtbar gemacht werden können (Graham et al., 1991). Latente Transitionsanalysen kombinieren insofern die Untersuchung latenter Strukturen, die üblicherweise anhand querschnittlicher Daten durchgeführt wird, mit längsschnittlichen Veränderungsmessungen (Nylund, 2007).

Nylund (2007) definiert latente Transitionsanalysen ferner als autoregressive Modelle. Autoregressive Modelle beschreiben anhand mehrerer Regressionen für zeitlich aufeinander folgende Datenpunkte die Beziehungen zwischen den zeitlich aneinandergrenzenden Outcomes (Nylund, 2007). Sie werden verwendet, wenn nicht von einer kontinuierlichen, sondern einer diskontinuierlichen Entwicklung auszugehen ist (Nylund, 2007). Ziel einer LTA ist es also, die diskontinuierlichen Entwicklungsverläufe verschiedener latenter Subgruppen in zeitlich aufeinanderfolgenden Stufen darzustellen (Graham et al., 1991). Die Untersuchung von stufen- oder phasenbezogenen Veränderungen eignet sich beispielsweise um Effekte von Interventionen auf die Entwicklung innerhalb unterschiedlicher Phasen zu bewerten (Graham et al., 1991). Im Kontext des hier vorgestellten Themas der beruflichen Orientierung könnte anhand latenter Transitionsanalysen beispielsweise die Wirkung einer zwischen zwei Messzeitpunkten besuchten Unterstützungsmaßnahme auf die Berufswahlkompetenzentwicklung analysiert werden.

Die Entwicklung der Individuen stellt eine dynamische latente Variable dar, die als abhängige Variable in den Analysen fungiert (Graham et al., 1991). Die Annahme dynamischer latenter Variablen – im Gegensatz zu statischen latenten Variablen – grenzt die LTA als autoregressives Modell von anderen latenten längsschnittlichen Analyseverfahren ab (Graham et al., 1991; Nylund, 2007).

Ein wesentlicher Vorteil latenter Transitionsanalysen im Vergleich zu anderen latenten Wachstumskurvenmodellen ist es, dass keine Restriktionen hinsichtlich der Gleichmäßigkeit der Abstände zwischen den Messzeitpunkten bestehen, sodass unterschiedliche Zeitsequenzen in die Analyse integriert werden können (Nylund, 2007). Die LTA lässt sich zu anderen längsschnittlichen latenten Analyseverfahren wie dem Latent Growth Curve Model (LGCM) außerdem wie folgt abgrenzen: Bei latenten Wachstumskurvenmodellen interessiert insbesondere die Form der Wachstumskurve, die die individuelle Veränderung beschreibt (Geiser, 2010). Ähnlich der LTA werden interindividuelle Unterschiede, also Unterschiede in den Wachstumskurven verschiedener Personen betrachtet, jedoch werden diese nicht zu Subgruppen zusammengefasst. Ein weiterer wesentlicher Unterschied besteht in der Annahme eines linearen bzw. kurvenlinearen und damit kontinuierlichen Wachstums bei LGCM (Geiser, 2010). Im Kontrast dazu ermöglicht die LTA, wie eingangs

bereits beschrieben, die Annahme eines zwischen den zeitlich aufeinander folgenden Stufen differierenden Wachstums (Graham et al., 1991). Anhand latenter Transitionsanalysen können also Phasen identifiziert werden, in denen sich für einzelne Subgruppen ein besonders starkes Wachstum manifestiert. Mit Blick auf die vier Entwicklungsphasen, die das Berufswahlkompetenzmodell von Driesel-Lange et al. (2010) postuliert, sowie der Annahme, dass die Entwicklung von Berufswahlkompetenz durch exogene Faktoren wie der sozialen Unterstützung durch das Umfeld (Mayhack & Kracke, 2010; Schindler, 2012) und durch externe Impulse beispielsweise durch das Absolvieren von Praktika, aber auch durch mögliche Rückschläge im Bewerbungsprozess geprägt wird, scheint die zeitlich abgegrenzte Betrachtung der Entwicklungsprozesse hier besonders relevant. Da im Rahmen der Berufswahlkompetenzentwicklung von einer individuellen, diskontinuierlichen Entwicklung ausgegangen wird, wurde die LTA anderen latenten Analysemodellen vorgezogen.

Vorgehensweise bei der Modellspezifizierung und Auswahl

Wie bereits für die LPA beschrieben, wird zu Beginn latenter Analysen das theoretische Konstrukt bestimmt, das abgebildet werden soll. In der klassischen LTA werden dieselben kategorialen Indikatoren über mehrere Zeitpunkte hinweg gemessen. Die Indikatoren repräsentieren also für unterschiedliche Messzeitpunkte jeweils das gleiche Konstrukt. Sie können sowohl dichotom als auch polytom ausgeprägt sein (Graham et al., 1991).

Nylund, Muthén, Nishina, Bellmore und Graham (2008) schlagen zur Spezifikation latenter Transitionsmodelle einen achtstufigen Prozess vor, über den auch Kovariaten und distale Outputvariablen in das Modell aufgenommen werden können. Dieser Leitfaden kann mit einem neueren Drei-Stufen-Verfahren von Asparouhov und Muthén (2014) kombiniert werden. Letzteres ermöglicht es, die Messparameter auch unter Einschluss von Kovariaten und distalen Outputvariablen stabil zu halten (Nylund-Gibson et al., 2014). Dieses Vorgehen ist besonders für die Verbindung von latenten Profilanalysen mit Wachstumsmodellen relevant (Asparouhov & Muthén, 2014). Die Vorteile der stabilen Messparameter finden jedoch auch Anwendung bei der längsschnittlichen Verbindung von latenten Profilanalysen wie in Kapitel 12. Im Folgenden werden nun die Schritte erläutert, die im Rahmen der Analysen in Kapitel 12 genutzt wurden.

Zunächst werden die manifesten Variablen anhand der deskriptiven Analysen auf fehlende Werte, Ausreißer und mögliche Kodierungsfehler hin untersucht (Nylund, 2007). Im nächsten Schritt werden anhand des Drei-Stufen-Verfahrens separat latente Profilanalysen für jeden Messzeitpunkt durchgeführt (Asparouhov & Muthén, 2014; Nylund-Gibson et al., 2014). Die vorerst getrennte Analyse der

einzelnen Messzeitpunkte ermöglicht es, die geeignete Profilanzahl für jeden Messzeitpunkt separat zu bestimmen und gleichzeitig mögliche Unterschiede zwischen den Profilmustern der Messzeitpunkte zu identifizieren (Nylund et al., 2008).

Im ersten Schritt wird eine Serie latenter Profilanalysen ohne Kovariaten durchgeführt, bei der beginnend mit einer einzigen Profilgruppe mit jeder Berechnung jeweils eine weitere Profilgruppe hinzugefügt wird. Die Auswahl des besten Modells und damit die Bestimmung der Profilanzahl erfolgt anhand von Fit-Indizes (Nylund et al., 2008, siehe dazu auch die Beschreibung der LPA in diesem Kapitel). Ist die Anzahl der Profile bestimmt, werden im zweiten Schritt die Kovariaten in das ausgewählte Modell als Ergänzungsvariablen eingeführt und für jeden Fall wird die Profilgruppe mit der jeweils höchsten Zugehörigkeitswahrscheinlichkeit gespeichert. Die Profile aus den Modellen mit und ohne Kovariaten werden verglichen, um die Stabilität der Gruppengrößen und den Einfluss der Kovariaten zu überprüfen (Nylund-Gibson et al., 2014). Das Modell im dritten Schritt verfügt nur noch über einen Indikator, die zugewiesene Profilgruppe aus dem ersten Schritt. Dabei werden die Schwellenwerte der latenten Profilvariable – des Indikators – nach der Berechnung von Asparouhov und Muthén (2014) fixiert (Nylund-Gibson et al., 2014). Die Verteilung der Stichprobe auf die Profilgruppen aus dem ersten Modellschritt wird mit der Verteilung aus dem dritten Modell verglichen, um Verschiebungen auszuschließen.

Sind für alle Messzeitpunkte die latenten Profilanalysen nach dem gerade beschriebenen dreistufigen Verfahren durchgeführt worden, können die Transitionen zwischen den Messzeitpunkten bereits anhand der in den querschnittlichen Profilanalysen zugewiesenen Profilen untersucht werden. Für jeden Transitionszeitpunkt werden die Wanderbewegungen zwischen den Profilgruppen in so genannten vorläufigen Transitionstabellen festgehalten. Bei einer LTA mit drei Messzeitpunkten bestehen folglich zwei Transitionszeitpunkte, vom ersten zum zweiten Messzeitpunkt sowie vom zweiten zum dritten Messzeitpunkt. Anhand dieser Tabellen lässt sich ablesen, wie viele Personen einer Profilgruppe in der gleichen Gruppe verblieben bzw. in eine andere Profilgruppe gewechselt sind (Nylund et al., 2008). Somit können erste Verlaufsmuster beobachtet werden. Im gleichen Schritt werden die Modelle auf Messinvarianzen hin überprüft (Nylund, 2007). Da in Kapitel 12 aufgrund theoretischer Überlegungen jedoch davon ausgegangen wurde, dass sich das Berufswahlkompetenzniveau der Gesamtstichprobe im zeitlichen Verlauf verändert, konnte auf die Testung von Messinvarianzen verzichtet werden (siehe auch Kia-Keating, Nylund-Gibson, Kia-Keating, Schock & Grimm, 2018).

Die Modellspezifizierung der LTA erfolgt anhand des gleichen Drei-Stufen-Prinzips wie für die zuvor beschriebenen Profilanalysen. In einer LTA mit drei

Messzeitpunkten wird jedem Fall für jeden der drei Messzeitpunkte jeweils eine Profilgruppe zugewiesen (Nylund, 2007). Im finalen Transitionsmodell werden diese latenten Variablen jeweils auf die latenten Variablen des vorherigen Messzeitpunkts regressiert. Die Werte zur Fixierung der Schwellen innerhalb der latenten Variablen werden aus den Berechnungen der einzelnen LPAs herangezogen (Nylund-Gibson et al., 2014). Analog zu latenten Profilanalysen werden den einzelnen Fällen als Ergebnis verschiedene, aus der latenten Transitionsanalyse hervorgehende Muster zugeordnet und als neue Ergebnisvariable zur anschließenden Interpretation gespeichert (Nylund, 2007).

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.