



**Universität
Zürich^{UZH}**

Analyse von hierarchischen Daten in R mittels Multilevel Analyse

Masterarbeit von

Noah Bosshart

Mat-Nr.: 13-747-141

Betreut durch

Prof. Dr. Carolin Strobl

3. Februar 2020

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
1 Abstract	V
2 Einleitung	1
3 Konzept und Anwendung von Multilevel Analyse	4
3.1 Beispiel zur Theorie	4
3.2 Intraklassen Korrelation	6
3.3 Lineare Modelle	8
3.3.1 Aggregation	9
3.3.2 Disaggregation	11
3.4 Hierarchische Lineare Modelle	13
3.4.1 <i>Random Intercept</i> Modell	14
3.4.2 <i>Random Intercept and Slope</i> Modell	17
3.5 Anwendung von Multilevel Analyse in R	17
3.5.1 R Pakete für die Multilevel Analyse	17
3.5.2 Aufbau eines Modells	17
3.5.3 Interpretation des Outputs	18
3.5.4 Vergleich von Hierarchischen Linearen Modellen	18
4 Simulationsstudie zur Multilevel Analyse	18
4.1 Herleitung der Forschungsfrage	18
4.2 Design der Simulationsstudie	19
4.2.1 Generierung von hierarchischen Daten	19
4.2.2 Manipulierte Faktoren	19
4.2.3 Konstante Faktoren	19

4.2.4	Untersuchte Faktoren	19
4.3	Ergebnisse der Simulationsstudie	19
5	Beschreibung und Anwendung der Shiny App	19
5.1	Was ist Shiny?	19
5.2	Ziel der Shiny App	19
5.3	Anwendung der Shiny App	19
6	Diskussion	19
7	Literaturverzeichnis	20
8	Anhang	21
A	R Code	21

Abbildungsverzeichnis

1	Zusammenhang zwischen der durchschnittlich gelösten Anzahl an Übungsaufgaben und der durchschnittlich erreichten Punktzahl pro Klasse	10
2	Zusammenhang zwischen der Anzahl gelöster Übungsaufgaben und erreichte Punktzahl mittels Disaggregation und Anwendung dieses Zusammenhangs auf jede der fünf Klassen	12
3	Zusammenhang zwischen der Anzahl gelöster Übungsaufgaben und der erreichten Punktzahl unter Berücksichtigung der Klassenzugehörigkeit	15
4	Residuenplot des linearen Modells und des <i>Random Intercept</i> Modells . . .	16

Tabellenverzeichnis

1	Beispiele für Level-1 und Level-2 Einheiten	2
2	Ausschnitt des simulierten Datensatzes	5
3	Mittlere Anzahl gelöster Übungsaufgaben und erreichte Punktzahl	9

1 Abstract

2 Einleitung

Hierarchische Daten treten häufig in den Sozialwissenschaften auf, unter anderem auch in der Psychologie (Snijders & Bosker, 2012). Von hierarchischen Daten wird gesprochen, wenn beispielsweise Daten von Schulkindern innerhalb verschiedener Schulklassen oder von Mitarbeitern aus mehreren Teams erhoben werden. Aber auch Daten aus Langzeitstudien werden als gruppiert bezeichnet, da mehrere Messzeitpunkte innerhalb einer Person gruppiert sind. Hierarchische Daten werden in Levels unterteilt, wobei Daten aus der niedrigsten Stufe als Level-1 Einheiten bezeichnet werden (Snijders & Bosker, 2012). Ein Beispiel für Level-1 Einheiten sind Schulkinder. Diese Schulkinder befinden sich wiederum in Klassen, die in der Hierarchiestufe höher sind und folglich als Level-2 Einheiten bezeichnet werden. Würde man nun in einer Studie nicht nur Schulkinder in Schulklassen, sondern auch die Schulen selbst berücksichtigen, würden die Schulen als Level-3 Einheit bezeichnet werden. Die Anzahl der Levels könnte man theoretisch beliebig hoch wählen, solange es das Studiendesign erlaubt und es aus der Perspektive der Forschungsfrage sinnvoll ist. Der Einfachheit halber beschränken wir uns im Laufe dieser Arbeit aber auf hierarchische Daten mit zwei Levels. In Tabelle 1 werden einige Beispiele für Level-1 und Level-2 Einheiten aufgeführt.

Dabei ist zu beachten, dass sich das Level der selben Einheit je nach Untersuchungsgegenstand ändern kann. Wie man in der Tabelle 1 erkennen kann, sind Familien einmal als Level-1 und einmal als Level-2 Einheit aufgeführt. Daher ist es wichtig die Level Bezeichnung nicht als starr zu betrachten. Vielmehr sollte man sich grundsätzlich an den niedrigsten Einheiten im Datensatz orientieren. Diesen Einheiten wird dann das Level-1 zugeschrieben.

In der Forschung ist es aus Kostengründen oder aus Gründen des Studiendesigns oft nicht möglich, solche gruppierte Datenstrukturen zu vermeiden (Snijders & Bosker, 2012; Woltman et al., 2012). Als eine von vielen Ursachen, die zur Entstehung solcher Datenstrukturen führt, nennen Snijders und Bosker (2012) *multistage sampling*. Unter *multistage sampling* versteht man, dass die Forschenden in der Datenerhebung auf in der Population vorhandene Gruppen zugreifen. Beispielsweise ist es Kostengünstiger zufällig 100 Schul-

Tabelle 1: Beispiele für Level-1 und Level-2 Einheiten

Level-1	Level-2
Schulkinder	Klasse
Studierende	Studienrichtungen
Kinder	Familien
Familien	Nachbarschaften
Mitarbeiter	Teams
Teams	Unternehmen
Patienten	Therapeuten
Therapeuten	Kliniken
Mehrere Messzeitpunkte	Person

kassen und von diesen Schulklassen wieder jeweils 10 Kinder auszuwählen als von 1000 Schulklassen jeweils nur einen Schulkind auszuwählen. Da man sonst in 1000 verschiedenen Schulklassen eine Studie durchführen müsste, um die gleiche Stichprobengrösse zu erreichen. Dieses Auswahlverfahren führt dazu, dass die erhobenen Daten nicht mehr voneinander unabhängig sind. Werden nun aus jeder Schulklasse 10 Schulkinder für eine Studie ausgewählt, ist es sehr wahrscheinlich, dass Schulkinder aus der selben Klasse zueinander ähnlichere Leistungen erzielen werden. Dieser Zusammenhang kann auf unterschiedliche Ursache zurückzuführen sein. Beispielsweise könnte die didaktischen Fähigkeiten der Lehrpersonen oder die Lichtverhältnisse im Klassenzimmer einen Einfluss auf die Leistungen der Kinder aus der selben Klasse haben. Das heisst, dass Einflussfaktoren aus unterschiedlichen Levels sich gegenseitig beeinflussen können.

Nach Snijders und Bosker (2012) gibt es unterschiedliche Formen, wie diese Einheiten zueinander in Beziehung stehen können. Ein Beispiel für einen Zusammenhang auf Level-1 wäre, dass die Lernmotivation eines Schulkindes sich auf seine Schulische Leistung auswirkt. Aber auch Level-2 Einheiten können sich gegenseitig beeinflussen. Das Klima

der Schulklasse könnte sich beispielsweise auf das Stressempfinden der Lehrperson auswirken. Hier wird von einem Zusammenhang innerhalb des Levels gesprochen, weil die unabhängige Variable (z.B. Lernmotivation, Klima der Schulklasse) auf dem gleichen Level wie die abhängige Variable (z.B. schulische Leistung, Stressempfinden) ist. Häufig ist es allerdings der Fall, dass es levelübergreifende Zusammenhänge zwischen den Einheiten gibt. So können beispielsweise die didaktischen Fähigkeiten einer Lehrperson (Level-2) und die Lernmotivation der Schulkinder (Level-1) die individuelle Leistung (Level-1) beeinflussen. Dieser Zusammenhang muss nicht zwingend direkt sein. Es kann auch vorkommen, dass die didaktischen Fähigkeiten den Zusammenhang zwischen Lernmotivation und individueller Leistung moderiert. In diesem Fall wird gemäss Snijders und Bosker (2012) von einer *cross-level interaction* gesprochen.

Werden diese Abhängigkeiten in der Analyse nicht berücksichtigt, kann dies zu einer erhöhten Fehler Typ-1 Rate führen (Dorman, 2008; McNeish, 2014). Das heisst, dass Forschende vermehrt zu Fehlschlüssen bezüglich des Einflusses ihrer Abhängigen Variablen gelangen und irrtümlich annehmen, einen Effekt eines Verfahrens gefunden zu haben, obwohl es diesen Effekt gar nicht gibt. Das Vorhandensein von hierarchischen Daten ist allerdings kein unlösbares Problem. Mit Analyseansätzen, die diese hierarchische Struktur der Daten berücksichtigen, lassen sich solche erhöhten Fehler Typ-1 Raten vermeiden. Einer dieser Ansätze ist die Multilevel Analyse, die im Fokus dieser Arbeit steht.

Diese Arbeit ist in drei Teile unterteilt. Im ersten Teil wird das Konzept und die Theorie der Multilevel Analyse behandelt. Dabei wird kurz auf die verschiedenen Methoden eingegangen, wie man Daten auf ihre hierarchische Struktur überprüfen kann. Anschliessend wird das zugrundeliegende statistische Modell der Multilevel Analyse vorgestellt und wie genau solche Modelle aufgebaut sind. Darauf folgend wird die Anwendung dieser Methoden in der Statistikumgebung R besprochen (R Core Team, 2019). Im zweiten Abschnitt dieser Arbeit wird eine Simulationsstudie durchgeführt, deren Ziel es ist, bereits vorhandene Ergebnisse in der Literatur zu replizieren und die Daseinsberechtigung der Multilevel Analyse von hierarchischen Daten zu festigen. Im dritten und letzten Abschnitt wird eine eigens programmierte Shiny Web-App vorgestellt (Chang et al., 2019), die zum einen das

Konzept der Multilevel Analyse visualisiert und dem Nutzer die Möglichkeit gibt, selbst die Simulationsstudie aus dem zweiten Abschnitt durchzuführen.

3 Konzept und Anwendung von Multilevel Analyse

Wie in der Einleitung erläutert wurde, gibt es viele Situationen in denen hierarchische Daten vorhanden sind und man zu Fehlschlüssen gelangen kann, wenn man diese Strukturen nicht berücksichtigt. In diesem Abschnitt wird nun etwas genauer auf das Konzept und die dahintersteckende Theorie der Multilevel Analyse eingegangen. Dazu wird zuerst ein simulierter Beispieldatensatz vorgestellt, anhand dessen die besprochenen Modelle erklärt werden. Als erstes wird auf die Probleme eingegangen, die durch die Verwendung von einfachen linearen Modellen entstehen. Anschliessend wird das hierarchische lineare Modell (HLM) als das zugrundeliegende statistische Modell der Multilevel Analyse eingeführt. Das HLM gilt als eine Erweiterung des einfachen linearen Modells Snijders & Bosker (2012). Dabei werden bei HLMs in *random intercept* und *random intercept and slope* Modelle unterschieden. Es werden beide Modellformen besprochen und dabei wird erläutert wie die beiden Faktoren Achsenabschnitt (engl. *intercept*) und Steigung (engl. *slope*) zusammenhängen. Nachdem die verschiedenen Formen von HLMs besprochen worden sind, wird in einem etwas praktischeren Teil die Anwendung von Multilevel Analyse in R anhand von Beispielen etwas näher gebracht.

3.1 Beispiel zur Theorie

In den folgenden Abschnitten wird die Theorie zur Analyse von hierarchischen Daten anhand eines Beispieldatensatzes erläutert. Bei dem Beispiel handelt es sich um insgesamt 150 Schulkindern aus 5 Schulklassen, die eine Mathematikprüfung geschrieben haben. Neben der erreichten Punktzahl wurde für jedes Kind zufällig ein Geschlecht, die Anzahl an gelösten Übungen, einen Wert für sozioökonomische Status und einen Intelligenzquotienten simuliert. Auf Stufe der Klasse wurden ausserdem noch die Anzahl Fenster im Klassenzim-

mer simuliert. Da dieser Datensatz selbst generiert wurde und aus keiner Studie entstammt, sollten Ergebnisse, die aus diesen Berechnungen entstehen nicht weiter interpretiert werden. Eine genaue Erläuterung wie dieser Datensatz generiert wurde, ist im Abschnitt über die Generierung von hierarchischen Daten zu finden. In Tabelle 2 sind zur Veranschaulichung die Daten von 10 Schulkindern aufgeführt.

Tabelle 2: Ausschnitt des simulierten Datensatzes

Schulkind Nr.	Klasse	Übungen	Punktzahl	Geschlecht	Anz. Fenster	SES	IQ
101	4	17	21	m	3	16	104
75	3	7	29	m	8	27	112
126	5	23	26	w	4	14	110
14	1	10	29	m	4	21	84
137	5	16	18	w	4	17	109
100	4	7	16	w	3	20	98
78	3	28	44	w	8	23	105
121	5	25	33	w	4	21	99
16	1	7	24	w	4	30	77
116	4	14	29	m	3	19	90

Betrachtet man die Variablen des Datensatzes, könnte man daraus schliessen, dass es sich um einen hierarchischen Datensatz mit zwei Levels handelt. Zu den Level-1 Variablen gehören alle Variablen die sich auf der Stufe der tiefsten Einheit (Schulkinder) befinden. Dazu zählen die Anzahl gelösten Übungen, die erreichte Punktzahl, das Geschlecht, der sozioökonomische Status und der IQ. Die beiden anderen Variablen Klasse und die Anzahl Fenster im Klassenzimmer gehören zur Level-2 Ebene. Um allerdings genau festzulegen, ob die hierarchische Struktur einen Einfluss auf die erreichte Punktzahl hat, benötigt es die Berechnung weiterer Kennwerte.

3.2 Intraklassen Korrelation

Der Einfluss einer hierarchischen Struktur auf eine abhängige Variable kann durch die Intraklassen Korrelation (IKK) beschrieben werden. Die Intraklassen Korrelation beschreibt den Grad der Ähnlichkeit von Level-1 Einheiten innerhalb einer Level-2 Einheit und kann als Verhältnis der Varianz zwischen den Level-2 Einheiten und der Gesamtvarianz beschrieben werden (Field et al., 2013; Snijders & Bosker, 2012; Twisk, 2006). Diese Varianzen ergeben sich gemäss Snijders und Bosker (2012) aus dem *random effects ANOVA* Modell, das bei der Modellierung von Multilevel Modellen oft auch als leeres Modell bezeichnet wird:

$$Y_{ij} = \mu + U_j + R_{ij} \quad (3.1)$$

Die abhängige Variable Y_{ij} beschreibt in unserem Beispiel die erreichte Punktzahl des Schulkindes i aus der Klasse j . Der Gesamtmittelwert aller Schulkinder wird mit μ bezeichnet, wobei U_j die zufällige Abweichung einer Klasse j und R_{ij} die zufällige Abweichung eines Schulkindes i der Klasse j von diesem Gesamtmittelwert beschreiben. Dabei ist zu beachten, dass der Erwartungswert beider Zufallsvariablen U_j und R_{ij} als 0 angenommen wird. Die Varianz von U_j wird als *between-group variance* τ^2 und von R_{ij} als *within-group variance* σ^2 bezeichnet.

Bei der IKK wird von einer Korrelation gesprochen, da es sich um die Korrelation zwischen zweier zufällig gewählter Level-1 Einheiten aus der selben Level-2 Einheit handelt. Bezogen auf unser Beispiel gibt die IKK an, wie stark sich Schulkinder aus der selben Klasse bezüglich ihrer erreichten Punktzahl ähneln. Ist die Korrelation zwischen den Schulkindern hoch, kann man davon ausgehen, dass die Klasse als Level-2 Einheit einen bedeutenden Anteil an der Gesamtvarianz erklärt. Ist die Korrelation niedrig hat die Klassenzugehörigkeit eher einen kleineren Einfluss auf die Prüfungsleistung. Dieser Zusammenhang wird etwas klarer, wenn man ihn anhand der Formel zur Berechnung der Intraklassen Korrelation Koeffizienten ρ_I erklärt:

$$\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2} \quad (3.2)$$

Dabei beschreibt τ^2 die *between-group variance*. In unserem Beispiel wäre das die Varianz der erreichten Punktzahl zwischen den verschiedenen Klassen. Die Gesamtvarianz setzt

sich aus der *between-group variance* und der *within-group variance* zusammen. Die Varianz innerhalb der Klassen wird, wie bereits erwähnt, mit σ^2 bezeichnet. Besteht nun innerhalb der Klassen eine kleine Varianz zwischen den Ergebnissen der Schulkinder ergibt sich eine grössere Intraklassen Korrelation. Steigt die Varianz innerhalb der Klassen an, wird der Nenner der Formel grösser und mit einem wachsenden Nenner, verringert sich die Intraklassen Korrelation.

Um nun zu überprüfen, ob in unserem Datensatz überhaupt abhängige hierarchische Strukturen vorhanden sind, können wir die IKK für unser Datensatz berechnen. Da die Populationswerte oft nicht bekannt sind, gibt es viele statistische Verfahren, um Schätzer für die nötigen Varianzen zu berechnen. Da diese Verfahren den Umfang dieser Arbeit sprengen würden und es viele Statistikprogramme gibt, die diese Berechnungen mit präzieseren Methoden durchführen können, werden in dieser Arbeit nur die computerbasierten Verfahren behandelt. Die restlichen Verfahren können aber in der gängigen Literatur zur Multilevel Analyse nachgeschlagen werden (z.B. Snijders & Bosker, 2012). Mit Hilfe des Statistikprogramms R wurden nun alle nötigen Varianzen geschätzt und in die Formel (3.2) eingesetzt¹:

$$\rho_I = \frac{12.33}{12.33 + 44} = 0.22 \quad (3.3)$$

Die daraus resultierende Intraklassen Korrelation von $\rho_I = 0.22$ weist darauf hin, dass 22% der Varianz in der erreichten Punktzahl in der Mathematikprüfung durch die Klassenzugehörigkeit erklärt wird. Gemäss Hedges und Hedberg (2007) werden in den Erziehungswissenschaften oft Intraklassen Korrelationen von 0.10 und 0.25 gefunden. Folglich liegt unsere IKK von $\rho_I = 0.22$ in einem realistischen Bereich. Eine Intraklassen Korrelation von $\rho_I > 0$ bedeutet aber noch nicht, dass eine Multilevel Analyse notwendig ist. Unter der Annahme, dass die zufällige Abweichungen der Schulkinder R_{ij} normalverteilt sind, kann gemäss Snijders und Bosker (2012) eine Varianzanalyse durchgeführt werden, um zu untersuchen, ob Gruppenunterschiede vorhanden sind. In unserem Fall führte die Varianzanalyse zu einem hoch signifikantem Ergebnis ($p < .001$) und es bestehen folglich Unterschiede zwischen den Klassen. Wir wissen nun nicht nur, wie viel Varianz durch die

¹Die Berechnung dieser Schätzer in R werden in Abschnitt 3.5 erläutert.

Klasse erklärt wird sondern auch, dass diese sich signifikant unterscheiden. Folglich sollten diese Daten mit einem Multilevel Ansatz analysiert werden.

3.3 Lineare Modelle

Bevor wir uns mit den hierarchischen linearen Modellen beschäftigen, werden die Grundlagen der linearen Regression kurz erläutert und aufgezeigt zu welchen Problemen es führen kann, wenn die hierarchische Datenstruktur ignoriert wird. Gemäss Gelman und Hill (2007) ist die lineare Regression eine Methode, die Veränderungen von Durchschnittswerten einer abhängigen Variablen durch eine lineare Funktion von Prädiktoren beschreibt. In etwas einfacheren Worten ausgedrückt, versucht die lineare Regression durch die Kombination von unabhängigen Variablen die mittlere Ausprägung einer abhängigen Variable zu beschreiben. Ein lineares Regressionsmodell kann wie folgt formuliert werden:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i, \text{ für } i = 1, \dots, n \text{ und } \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (3.4)$$

Dabei ist y_i die abhängige Variable von der Person i . In unserem Beispiel wäre das die erreichte Punktzahl des Schulkindes i . β_0 beschreibt den Achsenabschnittes (*intercept*) und ist die durchschnittlich erreichte Punktzahl in der Mathematikprüfung, wenn keine weitere Prädiktoren berücksichtigt werden. Die weiteren Regressionskoeffizienten β_1 bis β_k beschreiben für jede unabhängige Variable x_{i1} bis x_{ik} wie stark y_i des i -ten Schulkindes bei einer Zunahme um eine Einheit ansteigt. Die Regressionskoeffizienten β_1 bis β_k beschreiben also die Steigung (*slope*). Möchten wir in unserem Beispiel die erreichte Punktzahl durch die Anzahl gelöster Übungsaufgaben beschreiben, wäre x_{i1} die Anzahl gelöster Übungsaufgaben des i -ten Schulkindes und der dazugehörige Regressionskoeffizient β_1 gibt die Zunahme der Punktzahl in der Mathematikprüfung an. Der letzte Koeffizient des Regressionsmodells ist ϵ_i und wird als zufälliger Fehler oder Residuum bezeichnet. Das Residuum ist die normal verteilte zufällige Abweichung jedes i -ten Schulkindes, mit einem Erwartungswert von 0 und Varianz von σ^2 . Das bedeutet, dass es zwischen den Kindern zufällige Unterschiede in ihrer Prüfungsleistung gibt, die nicht durch das Regressionsmodell erfasst werden. Diese Unterschiede sind im Mittel aber 0.

Möchte man mit einem linearen Regressionsmodell die Daten unseres Beispiels untersuchen gibt es zwei Möglichkeiten. Die erste Möglichkeit ist die Aggregation, die häufig in den Sozialwissenschaften angewandt wird (Snijders & Bosker, 2012). Bei dieser Methode werden Mittelwerte für jede Klasse berechnet und anhand dieser wird dann ein lineares Modell erstellt. Die zweite Möglichkeit ist die Disaggregation, bei der die Klassenstruktur aufgelöst wird und alle 150 Schulkinder als unabhängige Werte in die Analyse einfließen.

3.3.1 Aggregation

Wie bereits erwähnt, werden bei der Aggregation für jede Level-2 Einheit Mittelwerte berechnet, die später in das Regressionsmodell einfließen. Ausgehend von unserem Beispiel könnte man sich nun für den Zusammenhang zwischen der Anzahl gelöster Übungsaufgaben und der erreichten Punktzahl in der Mathematikprüfung interessieren. In Tabelle 3 sind die relevanten Mittelwerte für jede der fünf Schulklassen aufgelistet.

Wird nun anhand dieser aggregierter Werte überprüft, wie genau die erreichte Punktzahl eines Schulkindes mit der Anzahl an gelösten Übungsaufgaben zusammenhängt, entstehen mehrere Probleme, die zu Verzerrungen und Fehlschlüssen führen können. Zum einen verändert sich die Forschungsfrage, da sich durch die Aggregation der Daten der Fokus von der Level-1 Ebene auf die Level-2 Ebene verschiebt (Snijders & Bosker, 2012; Woltman et al., 2012). Die abhängige Variable ist nun nicht mehr die erreichte Punktzahl jedes einzelnen Schulkindes, sondern die durchschnittlich erreichte Punktzahl einer Schulklasse.

Tabelle 3: Mittlere Anzahl gelöster Übungsaufgaben und erreichte Punktzahl

Klasse	Übungen	Punktzahl
1	13.1	21.5
2	12.8	29.3
3	13.5	30.7
4	15.7	25.6
5	17.5	24.7

Ein weiteres Problem ist der Verlust von Variabilität, die durch individuelle Unterschiede zwischen den Schulkindern entsteht. Dieser Verlust an Variabilität beträgt nach Raudenbush und Bryk 80-90% und kann zu massiven Fehlschlüssen über den Zusammenhang der Variablen führen (2002).

Betrachtet man die Regressionsgerade in Abbildung 1, sieht man, dass ein höhere Anzahl an gelöster Übungsaufgaben mit einer tieferen durchschnittlich erreichten Punktzahl zusammenhängt. Folglich könnte man daraus schliessen, dass dies auch auf Ebene der Schüler zutrifft und eine Erhöhte Anzahl an gelösten Übungsaufgaben mit einer tieferen Punktzahl in der Prüfung einhergeht. Diese Schlussfolgerung ist allerdings unzulässig, da man nicht von einer Korrelation zweier Level-2 Variablen auf den Zusammenhang von Level-1 Variablen schliessen darf (Snijders & Bosker, 2012). Diese fehlerhafte Schlussfolgerung wird auch als ökologischer Fehlschluss bezeichnet (Robinson, 2009).

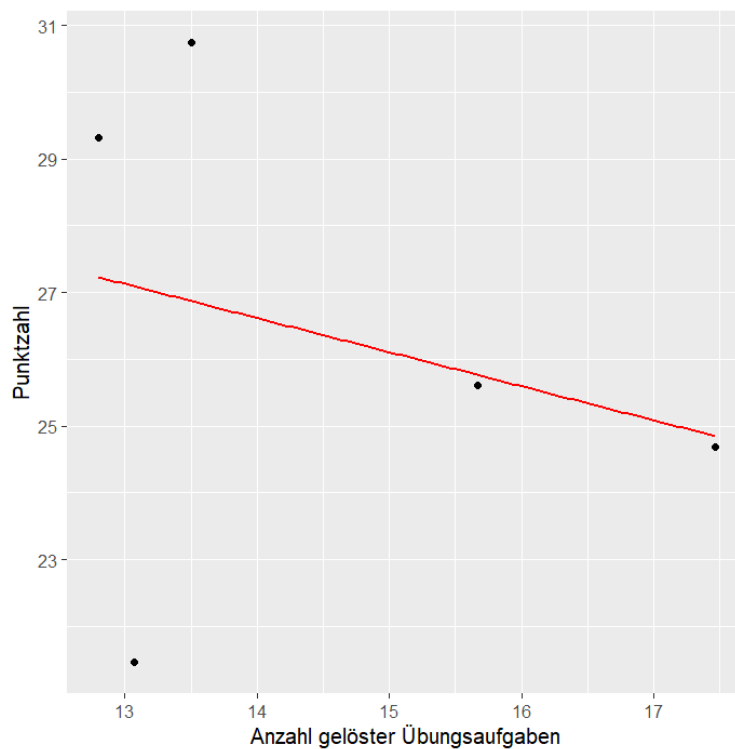


Abbildung 1: Zusammenhang zwischen der durchschnittlich gelösten Anzahl an Übungsaufgaben und der durchschnittlich erreichten Punktzahl pro Klasse

Die Analyse mittels Aggregation führt folglich nicht zu einem zufriedenstellenden Ergebnis und ist aufgrund der besprochenen Einschränkungen nicht geeignet, um Zusammenhänge auf Level-1 Ebene zu untersuchen.

3.3.2 Disaggregation

Die zweite Möglichkeit um hierarchische Daten mit einem linearen Regressionsmodell zu untersuchen ist die Disaggregation. Wie bereits angedeutet werden bei der Disaggregation alle Level-2 Variablen auf Level-1 Einheiten verteilt.

In unserem Beispiel werden also alle Schulkinder als von einander unabhängige Datenpunkte in die Analyse mit einbezogen. Dazu werden jedem Schulkind aus der selben Klasse die gleichen Werte der Level-2 Variablen zugeschrieben. In Tabelle 2 aus Abschnitt 3.1 kann man dieses Vorgehen bei den beiden Level-2 Variablen *Klasse* und *Fenster* beobachten. Durch diese Disaggregation von Level-2 Variablen auf Level-1 Einheiten werden Datensätze künstlich vergrößert und mögliche Variabilität, die zwischen den Level-2 Variablen besteht, wird ignoriert (Snijders & Bosker, 2012; Woltman et al., 2012). Folglich wird die geteilte Varianz zwischen Level-1 Einheiten nicht berücksichtigt und die Annahme, dass Fehler voneinander unabhängig sind, ist verletzt. Das führt dazu, dass die Effekte von Level-1 und Level-2 Variablen auf die abhängige Variable nicht voneinander getrennt werden können (Woltman et al., 2012). In unserem Beispiel würde das bedeuten, dass man den Einfluss der Anzahl an gelösten Übungsaufgaben nicht vom Einfluss der Klasse trennen kann. Ein weiteres Problem das durch Disaggregation entsteht, ist dass Abhängigkeiten innerhalb des Datensatzes unberücksichtigt bleiben (Woltman et al., 2012). Dies führt zu einer weiteren verletzten Annahme über die Unabhängigkeit von Beobachtungen. Die Verletzung dieser Annahme führt dazu, dass statistische Schätzer ungenau werden (Gelman & Hill, 2007; Snijders & Bosker, 2012; Woltman et al., 2012).

Auf der linken Seite der Abbildung 2 befindet sich die Regressionsgerade, die durch ein lineares Regressionsmodell entsteht, wenn man mit einem disaggregierten Datensatz arbeitet. Anhand dieser Regressionsgerade besteht ein positiver Zusammenhang zwischen der



Abbildung 2: Zusammenhang zwischen der Anzahl gelöster Übungsaufgaben und erreichte Punktzahl mittels Disaggregation und Anwendung dieses Zusammenhangs auf jede der fünf Klassen

Anzahl gelöster Übungsaufgaben und der erreichten Punktzahl in der Mathematikprüfung, so dass die erreichte Punktzahl mit steigender Anzahl an gelöster Übungsaufgaben zunimmt. Wie vorhin bereits erwähnt, wird in dieser Analyse aber nicht berücksichtigt, dass die Schulklasse selbst einen Effekt auf die erreichte Punktzahl haben kann. Dieser Effekt wird klar, wenn man die rechte Seite der Abbildung 2 betrachtet. Für jede der fünf Klassen wurde die selbe Regressionsgerade, die aus dem disaggregierten Datensatz entsteht, über die Daten gelegt. Man kann relativ einfach erkennen, dass es gewisse Klassen gibt, bei denen mehr Schulkinder über oder unter der Regressionsgerade liegen. Des weiteren kann man erkennen, dass es nicht optimal ist, wenn für alle Klassen die selbe Steigung der Regressionsgerade verwendet wird. Betrachten wir beispielsweise die zweite Klasse, kann man erkennen, dass diese Schulkinder einen viel stärkeren Zusammenhang zwischen gelösten Übungsaufgaben und erreichter Punktzahl verzeichnen als die erste Klasse. Man könnte nun mit Hilfe einer Dummy-Kodierung den Einfluss von Klassen berücksichtigen, dazu müsste aber für jede Klasse einen zusätzlichen Parameter in das Modell aufgenommen

werden. Da es grundsätzlich erstrebenswert ist, möglichst sparsame Modelle zu bilden ist auch dies keine optimale Lösung.

Die Aggregation als auch die Disaggregation der Daten unterliegen massiven Einschränkungen und führen zu keinem zufriedenstellenden Ergebnis. Es erfordert folglich ein weiteres Modell, das Zusammenhänge innerhalb und zwischen Level-2 Einheiten abbilden kann ohne sich dabei auf eine Analyseinheit festzulegen.

3.4 Hierarchische Linearen Modelle

In den letzten Abschnitten wurde angenommen, dass sich die Regressionskoeffizienten β_0 und β_1 feste Werte sind, die sich nicht verändern. In Abbildung 2 aus dem vorherigen Abschnitt konnte man aber erkennen, dass diese Annahme nicht in allen Fällen zu einem erwünschten Ergebnis führt. In unserem Beispiel gibt es offensichtlich Klassen, die eine über- oder unterdurchschnittliche erreichte Punktzahl verzeichnen. Man kann nun annehmen, dass diese Regressionskoeffizienten zufällig sind. In diesem Kontext versteht man unter zufällig aber nicht, dass die Koeffizienten irgendwie gewählt werden können, sondern vielmehr, dass diese Koeffizienten variieren können.

In den folgenden Abschnitten werden nun hierarchische lineare Modelle besprochen, mit denen es möglich ist solche zufällige Koeffizienten zu schätzen. Als erstes wird das *Random Intercept* Modell vorgestellt. Dieses Modell geht davon aus, dass die Höhe des Achsenabschnittes (*intercept*) von der Gruppenzugehörigkeit abhängt und schätzt folglich mehrere verschiedene Achsenabschnitte. Das zweite besprochene Modell ist das *Random Intercept and Slope* Modell, bei dem sich nicht nur der Achsenabschnitt, sondern auch die Steigung (*slope*) in Abhängigkeit der Gruppe unterscheidet. Dabei wird der Fokus vor allem darauf gesetzt, dass das Konzept von hierarchischen linearen Modellen verstanden wird. Wie genau solche HLMs berechnet und analysiert werden, wird im Abschnitt 3.5 behandelt.

3.4.1 *Random Intercept* Modell

Das *Random Intercept* Modell ermöglicht es für jede Gruppe unterschiedliche Achsenabschnitte zu schätzen. Die einfachste Form eines *Random Intercept* Modells ist ein Modell, das nur den Koeffizienten für den Achsenabschnitt β_{0j} und das Residuum ϵ_i enthält. Dieses Modell wird wie folgt beschrieben:

$$\begin{aligned}\text{Level 1:} \quad y_{ji} &= \beta_{0j} + \epsilon_i \\ \text{Level 2:} \quad \beta_{0j} &= \gamma_{00} + U_{0j}\end{aligned}\tag{3.5}$$

Bei dieser Darstellung handelt es sich um die hierarchische Notation der Gleichung, da die einzelnen Gleichungen gleich dem dazugehörigen Level zugeordnet werden. Dies wird klarer, wenn man es in Bezug zu unserem Beispiel betrachtet. Auf Level-1 befindet sich die Regressionsgleichung für die erreichte Punktzahl jedes einzelnen Schulkindes i aus der Klasse j . Dabei kann man erkennen, dass der Regressionskoeffizient β_{0j} von der Klasse j abhängt und folglich für jede Klasse einen anderen Wert einnimmt. Da die Klasse eine Level-2 Variable ist, befindet sich die Gleichung für β_{0j} auf Level-2. Dabei ist γ_{00} der Gesamtmittelwert und U_{0j} die zufällige Abweichung der Klasse j vom Gesamtmittelwert. Substituiert man die Gleichung von Level-2 in die Gleichung von Level-1 gelangt man zur flachen Notation dieses *Random Intercept* Modells:

$$y_{ji} = \gamma_{00} + U_{0j} + \epsilon_i\tag{3.6}$$

Diese Gleichung entspricht dem leeren Modell aus Abschnitt 3.2, anhand dessen man die Intraklassen Korrelation berechnet. Ähnlich wie bei der normalen linearen Regression können diesem Modell nun weitere Variablen hinzugefügt werden, um die Varianz in der erreichten Punktzahl des Schulkindes i aus der Klasse j zu erklären. In unserem Beispiel ergänzen wir das Modell mit nur einer weiteren Variable x_{ij} , die der Anzahl gelöster Übungsaufgaben entspricht:

$$\begin{aligned}\text{Level 1:} \quad y_{ji} &= \beta_{0j} + \beta_1 x_{ij} + \epsilon_i \\ \text{Level 2:} \quad \beta_{0j} &= \gamma_{00} + U_{0j} \\ \beta_1 &= \gamma_{10}\end{aligned}\tag{3.7}$$

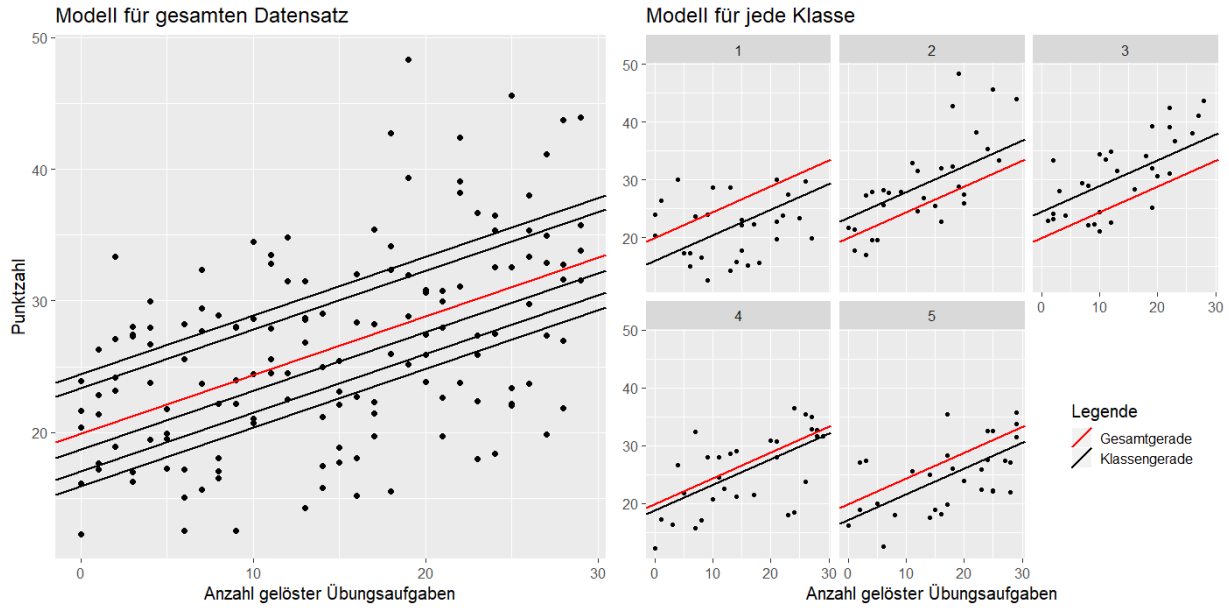


Abbildung 3: Zusammenhang zwischen der Anzahl gelöster Übungsaufgaben und der erreichten Punktzahl unter Berücksichtigung der Klassenzugehörigkeit

Da es sich hier um ein *Random Intercept* Modell handelt, bleibt die Steigung für alle Klassen gleich. Dies kann man an der Gleichung des Koeffizienten β_1 erkennen, da es keine zufällige Abweichung in Abhängigkeit der Klasse j von der Gesamtsteigung γ_{10} gibt. Werden nun die beiden Gleichungen auf Level-2 in die Gleichung auf Level-1 eingesetzt, gelangen wir wieder zur flachen Notation des *Random Intercept* Modells:

$$y_{ji} = \gamma_{00} + U_{0j} + \gamma_{10}x_{ij} + \epsilon_i \quad (3.8)$$

In Abbildung 3 kann man nun die Graden für jede einzelne Klasse erkennen. Die rote Gerade entspricht der linearen Regressionsgerade, die durch die Disaggregation entsteht und wird hier als Vergleichswert zu den anderen Geraden verwendet. Auf der linken Seite der Abbildung erkennt man relativ schnell, dass es bedeutende Unterschiede zwischen den Klassen gibt. Werden die Geraden für jede einzelne Klasse betrachtet, erhält man einen Überblick über das Leistungslevel der verschiedenen Klassen. Beispielsweise kann man erkennen, dass die Klassen eins und fünf eher tiefere und die Klassen zwei und drei eher höhere Punktzahlen erreichen. Ebenfalls kann man beobachten, dass die Geraden des

hierarchischen linearen Modells besser zu den Daten passen. Dies wird etwas klarer, wenn man die Residuenplots des *Random Intercept* Modells und des normalen linearen Modells vergleicht. In Abbildung 4 sind die beiden Residuenplots abgebildet. Auch wenn die Unterschiede nicht all zu gross sind, kann man erkennen, dass es beim *Random Intercept* Modell weniger grosse Abweichungen gibt zwischen den tatsächlich beobachteten Werten und den vom Regressionsmodell erwarteten Werten. Allerdings fällt beim Residuenplot des *Random Intercept* Modells auf, dass an den Endpunkten die Residuen nicht mehr um den Nullpunkt verteilt sind, beim linearen Modell allerdings schon. Die Ursache dafür besteht darin, dass die Residuen nun für jede einzelne Regressionsgerade der Klassen berechnet werden. Da bei dem normalen linearen Modell die Residuen alle nur an einer Geraden berechnet werden, gleichen sich die Residuen der Schulkinder aus leistungstärkeren Klassen mit den Kindern aus leistungsschwächeren Klassen aus. Betrachtet man in Abbildung 3 die rechte Seite, passen die Regressionsgeraden des *Random Intercept* Modells besser zu den Daten als das lineare Modell. Allerdings gibt es immer noch Schulkinder, die noch nicht optimal durch die Gerade beschrieben werden. Bei der zweiten und dritten Klasse erreichten beispiels-

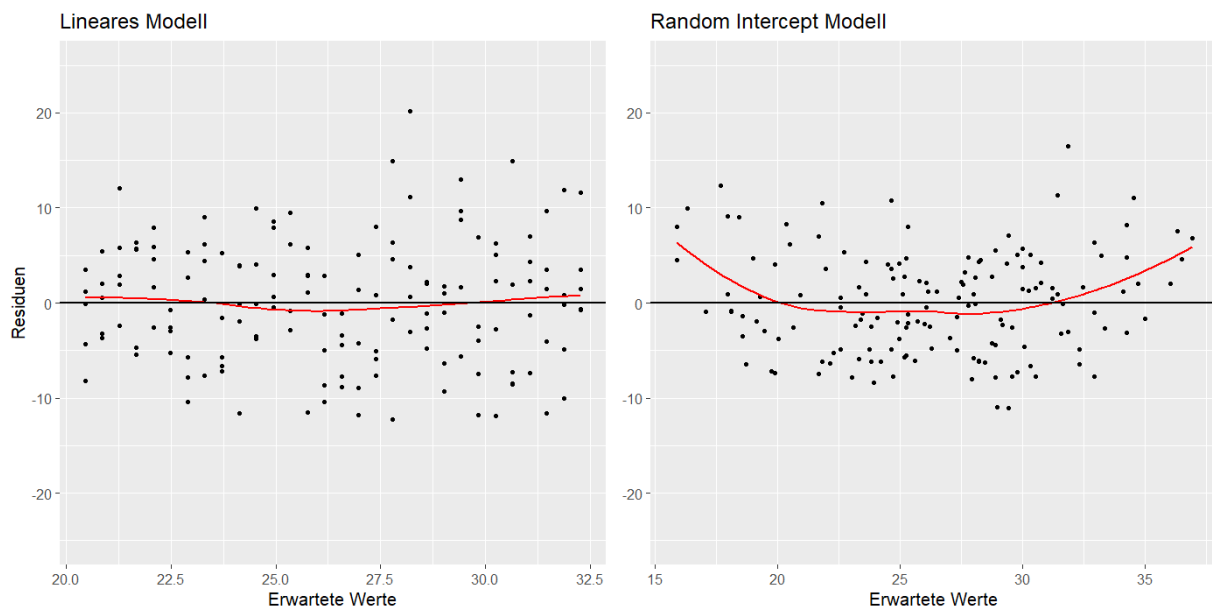


Abbildung 4: Residuenplot des linearen Modells und des *Random Intercept* Modells

weise Schulkinder, die viele Übungen gelöst haben eine noch viel höhere Punktzahl als vom Modell angenommen wird. Diese Ungenauigkeit führt folglich zu einem Anstieg der Residuen an den Endpunkten. Diese Ungenauigkeit an den Endpunkten kann aber auch als Information genutzt werden. Anscheinend gibt es in unserem Datensatz Klassen, bei denen die Schulkinder einen stärkeren oder schwächeren Anstieg der erreichten Punktzahl bei steigender Anzahl gelöster Übungsaufgaben verzeichnen. Daraus könnte man nun folgern, dass sich nicht nur der Achsenabschnitt zwischen den Klassen unterscheidet, sondern auch die Steigung.

3.4.2 *Random Intercept and Slope* Modell

3.5 Anwendung von Multilevel Analyse in R

Da nun das Konzept der hierarchischen linearen Modellen erläutert wurde, gelangen wir hier in diesem abschnitt zu einer etwas mehr anwendungsbezogenene dinge. Dabe möchte ich hier besprechen, wie nun solche HLMS in er erstellt und analysiert werden. Dazu erkläre ich zuerst welche Pakete es gibt um HLMS zu erstellen und welches hier genutzt wird. Ausserdem wird gezeigt, wie modelle miteinander verglichen werden können und welche Varianzen genau durch welche Variablen erklärt werden.

3.5.1 R Pakete für die Multilevel Analyse

Beschreibung von lme4 und grund warum in dieser Arbeit nur mit diesem Paket gearbeitet wird. (Buch und Studie von D. Bates)

3.5.2 Aufbau eines Modells

Die meisten Modelle erlauben nicht mehr als 2-3 Random Slopes und konvergieren nicht (Snijders & Bosker, 2012)

3.5.3 Interpretation des Outputs

3.5.4 Vergleich von Hierarchischen Linearen Modellen

Modelle welche sich nur in fixen Effekten unterscheiden sollten mit ML und Modelle welche sich in zufälligen Effekten unterscheiden mit REML verglichen werden Snijders & Bosker (2012)

Tests für feste Effekte Wald-Test Snijders & Bosker (2012) Inkl. Dummy-Test

Deviance Tests ebenfalls verwendbar für feste Effekte. Bei Random Intercept an chi-square verteilung mit $df = \text{anz. veränderte variable teile}$ (wichtig fixed effect müssen gleich bleiben, wenn mit REML, sonst ML)

Da Varianzen nicht negativ werden können, wird oft einseitig getestet. Konservativere Möglichkeit durch halbierung des testwertes (SZweiseitiges Testen”).

Deviance Tests für Random Slope etwas aufwändiger, $df = m1 - m0 = p + 1$ (anz. covarianzen p , von denen sich das $m0$ zu $m1$ unterscheiden $+ 1$ varianz) Prüfwert wird für $df = p$ und für $df = p+1$ in einer chi-quadrat verteilung bestimmt. danach mittelwert davon ergibt den eigentlichen prüfwert.

Konfidenzintervall am besten durch profile likelihood (via lme4 Paket). Profile likelihood verhindert, dass Konfidenzintervalle den Wert 0 Unterschreiten, da Varianzen nicht negativ sein können.

Wenn diese Methode nicht vorhanden ist können andere Methoden gewählt werden, die allerdings nicht so genau/reliabel sind.

Proportionale Reduktion der Varianz und Pseude R Squared (Zitation nötig!)

4 Simulationsstudie zur Multilevel Analyse

4.1 Herleitung der Forschungsfrage

Es gibt schon Tutorials etc. wie man HLM in der Forschung einsetzt. Dabei achten auf Kennwerte (ICC und DEFF). Studien haben gezeigt, dass Fehler Typ-1 Rate steigt wenn

MLM anstatt HLM) Studien zitieren, Dorman, Neith, Etc. Es stellt sich aber auch die Frage, wie es genau mit Treatments aussieht (studie treatment zitieren) auch diese haben einen erhöhte Typ-1 Rate gefunden. Ziel: replikation der ergebnisse, dass Fehler typ-1 rate erhöht ist und in einen für psychologiestudenten relevanten kontext bringen um das Konzept der HLM den studierenden zu verkaufen. H1: betas werden genau geschätzt H2: SE bei Effekt von 0 zu klein bei MLM -> folglich zu viele p-werte > 0.05

4.2 Design der Simulationsstudie

4.2.1 Generierung von hierarchischen Daten

4.2.2 Manipulierte Faktoren

4.2.3 Konstante Faktoren

4.2.4 Untersuchte Faktoren

4.3 Ergebnisse der Simulationsstudie

5 Beschreibung und Anwendung der Shiny App

5.1 Was ist Shiny?

5.2 Ziel der Shiny App

5.3 Anwendung der Shiny App

6 Diskussion

7 Literaturverzeichnis

- Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. (2019). shiny: Web Application Framework for R [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=shiny> (R package version 1.3.2)
- Dorman, J. P. (2008). The effect of clustering on statistical tests: an illustration using classroom environment data. *Educational Psychology*, 28 (5), 583–595.
- Field, A., Miles, J. & Field, Z. (2013). *Discovering statistics using r* (Reprinted Aufl.). Los Angeles: Sage.
- Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. United Kingdom: Cambridge University Press. (Includes bibliographical references (pages 575-600) and indexes)
- Hedges, L. V. & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29 (1), 60-87. Zugriff auf <https://doi.org/10.3102/0162373707299706> doi: 10.3102/0162373707299706
- McNeish, D. M. (2014). Analyzing clustered data with ols regression: The effect of a hierarchical data structure. *Multiple Linear Regression Viewpoints*, 40 (1), 11–16.
- R Core Team. (2019). R: A language and environment for statistical computing [Software-Handbuch]. Vienna, Austria. Zugriff auf <https://www.R-project.org/>
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Bd. 1). Sage.
- Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International journal of epidemiology*, 38 (2), 337–341.
- Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2. Aufl.). Los Angeles: SAGE.

Twisk, J. W. R. (2006). *Applied multilevel analysis: A practical guide for medical researchers*. Cambridge University Press. doi: 10.1017/CBO9780511610806

Woltman, H., Feldstain, A., MacKay, J. C. & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology*, 8 (1), 52–69.

8 Anhang

A R Code