# Analyse von hierarchischen Daten in R mittels Multilevel Analyse

Masterarbeit von

Noah Bosshart

Mat-Nr.: 13-747-141

Betreut durch

Prof. Dr. Carolin Strobl

22. Januar 2020

# Inhaltsverzeichnis

# Abbildungsverzeichnis

# Tabellenverzeichnis

# 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 betrachtet. 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 Verfahren 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 Multlilevel 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 da Daseinsberechtigung von Mulitlevel 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 wenn diese Strukturen nicht berücksichtigt werden, kann man zu Fehlschlüssen gelangen. 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 dieses Beispiels werden Probleme aufgezeigt, die entstehen, wenn man einfache lineare Modelle verwendet und anschliessend wird das hierarchische lineare Modell (HLM) eingeführt. Das HLM ist das zugrundeliegendes statistisches Modell der Multilevel Analyse und 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. und 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 Klassenzimmer 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 Ausprägung einzelner Variablen des Datensatzes, kann man erkennen, dass gewisse Variablen sich nicht zwischen den Schulkindern verändern. Wenn sich Variablen über einzelne Beobachtungen hinweg nicht verändern, kann das ein Hinweis dafür sein, dass es sich um hierarchische Daten handelt. Unser Datensatz hat in der Tat ein hierarchische Struktur mit zwei Levels. 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 zu Level-2.

## 3.2 Lineare Modelle

Bevor wir uns mit den hierarchischen linearen Modellen beschäftigen, werden hier noch einmal kurz die Grundlagen der linearen Regression erläutert. 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:

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{ik} X_{ik} + \epsilon_i, \text{ für } i = 1, \ldots, n \text{ und } \epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{3.1}$$

Dabei ist $y_i$ die abhängige Variable von der Person $i$. In unserem Beispiel wäre das die erreichte Punktzahl des Schulkindes $i$. $\beta_0$ beschreibt den Achsenabschnittes (*intercept*) und ist die durchschnittlich erreichte Punktzahl in der Mathematikprüfung, wenn keine weitere Prädiktoren berücksichtigt werden. Die weiteren Regressionskoeffizienten $\beta_1$ bis $\beta_k$ 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 $\beta_1$ bis $\beta_k$ 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 $\beta_1$ gibt die Zunahme der Punktzahl in der Mathematikprüfung an. Der letzte Koeffizient des Regressionsmodells ist $\epsilon_i$ 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 $\sigma^2$. 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, bei der Mittelwerte für jede Klasse berechnet werden und die häufig in den Sozialwissenschaften angewandt wird (Snijders & Bosker, 2012). Die zweite Möglichkeit ist die Disaggregation, bei der die Klassenstruktur aufgelöst wird und alle 150 Schulkinder als unabhängige Werte in die Analyse einfliessen.

### 3.2.1 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. 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).

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 |

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).

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.

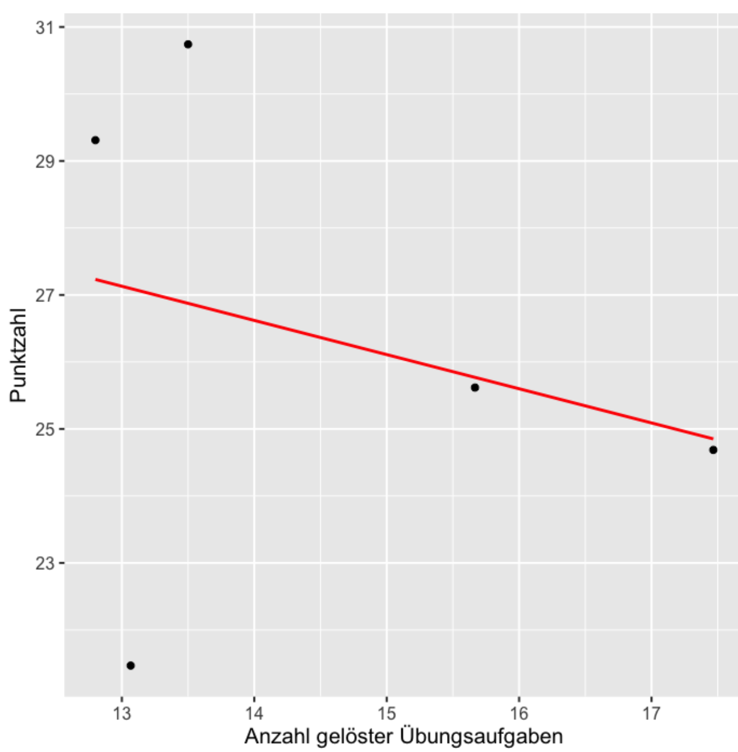


Abbildung 1: Zusammenhang zwischen der durchschnittlich gelösten Anzahl an Übungsaufgaben und der durschnittlich erreichten Punktzahl pro Klasse

### 3.2.2 Disaggregation

Die zweite Möglichkeit um hierarchsiche 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 geteilt 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 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 erreicht Punktzahl haben kann. Dieser Effekt

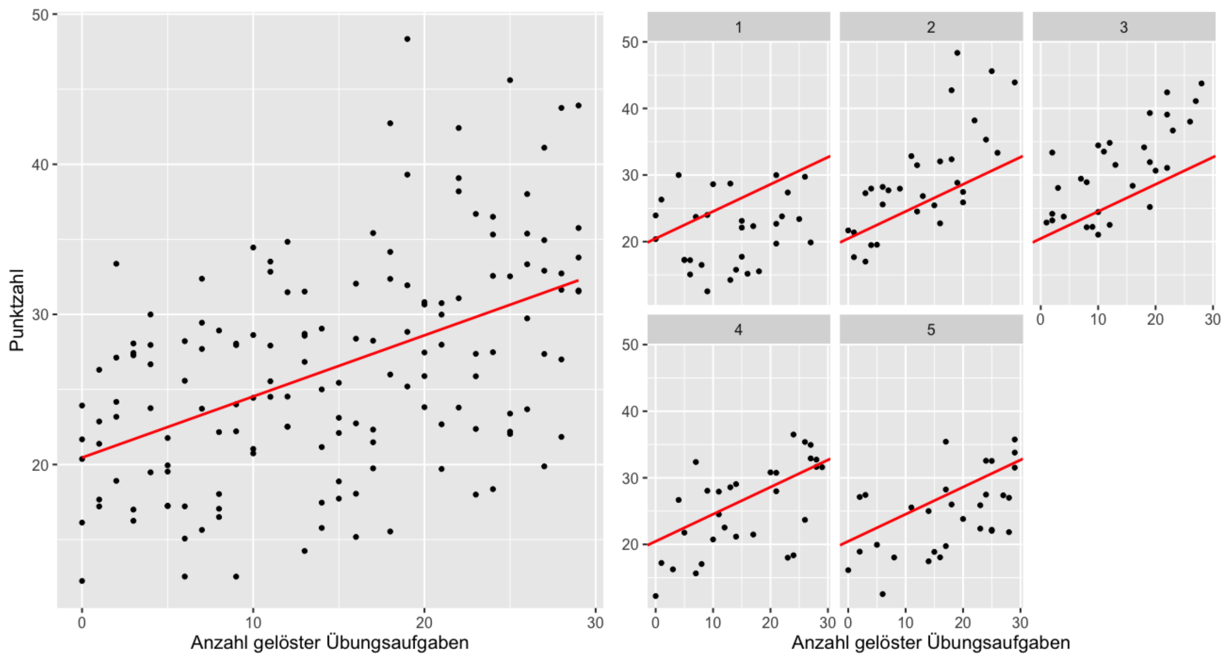


Abbildung 2: Zusammenhang zwischen der Anzahl gelöster Übungsaufgaben und erreichte Punktzahl mittels Disaggregation und Anwendung dieses Zusammenhangs auf jede der fünf Klassen

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 Schüler 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.

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.

## 3.3 Hierarchische Linearen Modelle

Hierarchische Lineare Modelle sind, wie bereits angedeutet eine Erweiterung von einfachen linearen Regressionsmodellen.

### 3.3.1 Intraklassen Korrelation

Besprechen von ICC und Design Effekt (Vlg. Dazu Guide ML Analysis von J. Peugh 2009)

### 3.3.2 *Random Intercept* Modell

### 3.3.3 *Random Intercept and Slope* Modell

## 3.4 Anwendung von Multilevel Analyse in R

### 3.4.1 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)

### 3.4.2 Aufbau eines Modells

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

### 3.4.3 Interpretation des Outputs

### 3.4.4 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 = 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 druch 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.

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)

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.

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