



**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

15. Januar 2020

Inhaltsverzeichnis

Abbildungsverzeichnis	4
Tabellenverzeichnis	5
1 Abstract	6
2 Einleitung	7
3 Konzept und Anwendung von Multilevel Analyse	10
3.1 Simuliertes Beispiel	10
3.2 Einfache Lineare Modelle	10
3.3 Hierarchische Linearen Modelle	10
3.3.1 Intraklassen Korrelation	11
3.3.2 <i>Random Intercept</i> Modell	11
3.3.3 <i>Random Intercept and Slope</i> Modell	11
3.4 Anwendung von Multilevel Analyse in R	11
3.4.1 R Pakete für die Multilevel Analyse	11
3.4.2 Aufbau eines Modells	11
3.4.3 Interpretation des Outputs	11
3.4.4 Vergleich von Hierarchischen Linearen Modellen	11
4 Simulationsstudie zur Multilevel Analyse	12
4.1 Herleitung der Forschungsfrage	12
4.2 Design der Simulationsstudie	13
4.2.1 Generierung von hierarchischen Daten	13
4.2.2 Manipulierte Faktoren	13
4.2.3 Konstante Faktoren	13
4.2.4 Untersuchte Faktoren	13
4.3 Ergebnisse der Simulationsstudie	13

5	Beschreibung und Anwendung der Shiny App	13
5.1	Was ist Shiny?	13
5.2	Ziel der Shiny App	13
5.3	Anwendung der Shiny App	13
6	Diskussion	13
7	Literaturverzeichnis	14
8	Anhang	14
A	R Code	14

Abbildungsverzeichnis

Tabellenverzeichnis

1	Beispiele für Level-1 und Level-2 Einheiten	7
---	-------------------------------------------------------	---

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.

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

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

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 Schulkassen und von diesen Schulkassen wieder jeweils 10 Kinder auszuwählen als von 1000 Schulkassen 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 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 da Daseinsberechtigung von 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 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 zuerst Probleme aufgezeigt, die entstehen, wenn man einfache lineare Modelle verwendet und anschliessend wird das hierarchische lineare Modell (HLM) eingeführt. Das HLM ist das zugrundeliegende statistische 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 *intercept* und *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 Simuliertes Beispiel

In unserem Beispiel handelt es sich um eine Forschungsgruppe, die untersuchen möchte wie stark der Einfluss der Anzahl Mathematikektionen in einer Woche auf die Leistungen in einer Mathematikprüfung ist. Dafür wurden aus 10 Schulklassen jeweils 10 Schüler ausgewählt

3.2 Einfache Lineare Modelle

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 $df = \text{anz. veränderte variable teile}$ (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 Simulationstudie 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.
- 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/>
- 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