



**Universität
Zürich^{UZH}**

Multilevel Analyse in R

Interaktive Einführung in Theorie und Anwendung

Masterarbeit von

Noah Bosshart

Betreut durch

Prof. Dr. Carolin Strobl

December 9, 2019

Contents

1	Abstract	3
2	Einleitung	4
2.1	Genestete Datenstrukturen	4
2.2	Problematik von linearen Modellen	4
3	Multilevel Analyse von genesteten Daten in R	5
3.1	R Pakete für die Multilevel Analyse	5
3.2	Wann werden Multilevel Modelle eingesetzt?	5
3.3	Aufbau von Multilevel Modellen	5
3.3.1	Das Null Modell	6
3.3.2	Das Level-1 Modell	6
3.3.3	Das Level-2 Modell	6
3.3.4	Cross-Level Interaktion	6
3.4	Vergleich von Multilevel Modellen	6
3.5	Kennwerte von Multilevel Modellen	7
4	Multilevel Analyse mit Shiny	7
5	Diskussion	7
6	References	8
7	Abbildungsverzeichnis	8
8	Anhang	8

1 Abstract

2 Einleitung

Genestete Datenstrukturen findet man in vielen Aspekten unseres Lebens, sei es Schüler in Klassen, Teams in Organisationen, Kinder in Familien oder Messungen von Längsschnittstudien. In diesen Datenstrukturen bestehen gewissen Abhängigkeiten zwischen den einzelnen Messeinheiten. Das bedeutet, dass beispielsweise Messungen innerhalb einer Klasse höhere Korrelationen aufweisen als Messungen zwischen den Klassen. Etwas einfacher ausgedrückt, werden Schüler aus der selben Klasse zueinander ähnlicher sein als zu Schülern aus anderen Klassen. Diese Gegebenheit kann auf viele verschiedene Ursachen zurückzuführen sein, wie zum Beispiel die Fähigkeiten der Lehrperson oder die Qualität der Lehrmaterialien.

Würden Daten mit diesen Strukturen mit einem linearen Regressionsmodell analysiert werden, könnte das zu fehlerhaften Ergebnissen führen, da diese Form der Anal Diese Arbeit befasst sich nun damit, wie solche Datenstrukturen mittels Multilevel Analyse berücksichtigt werden können, um Fehlschlüsse zu vermeiden.

Bevor wir uns aber mit den theoretischen Grundlagen der Multilevel Analyse befassen können, muss geklärt werden, wie genau solche genesteten Datenstrukturen aufgebaut sind. Dazu werden im folgenden Abschnitt genestete Datenstrukturen genauer beschrieben und es wird beschrieben, wie man Daten mit solchen Strukturen in der Statistiksoftware R simulieren kann (R Core Team, 2019).

2.1 Genestete Datenstrukturen

2.2 Problematik von linearen Modellen

Was passiert wenn genestete Strukturen ignoriert (aggregiert) werden (Snijders & Bosker, 2012).

Stichproben sollten immer zufällig gezogen werden, dies ist häufig aber nicht der Fall, da es aus Kostengründen einfacher ist bereits vorhandene Gruppen (Cluster) zu ziehen. Beispielsweise sind das Klassen, Teams, Nachbarschaften, etc. Sobald aber solche Cluster gezogen werden, bestehen Abhängigkeiten zwischen den einzelnen Datenpunkte innerhalb

der Cluster. Folglich ist die Annahme der Unabhängigkeit der Varianzen von linearen Modellen verletzt.

Bei steigender Intraklassenkorrelation nimmt ebenfalls der α -Fehler zu (Dorman, 2008).

3 Multilevel Analyse von genesteten Daten in R

3.1 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)

3.2 Wann werden Multilevel Modelle eingesetzt?

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

3.3 Aufbau von Multilevel Modellen

Aufbau erklären. Was ist das richtige Vorgehen um ein Multilevel Modell zu erstellen. Nullmodell bis hin zu Cross-Level Modellen etc. An Guides zu Multi Level Modellen Orienteieren! (Snijders & Bosker, 2012) (Weitere Guides / Tutorials zu MLM finden)

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

3.3.1 Das Null Modell

3.3.2 Das Level-1 Modell

3.3.3 Das Level-2 Modell

3.3.4 Cross-Level Interaktion

3.4 Vergleich von Multilevel 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 ("Zweiseitiges 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.

3.5 Kennwerte von Multilevel Modellen

4 Multilevel Analyse mit Shiny

5 Diskussion

6 References

Dorman, J. P. (2008). The effect of clustering on statistical tests: an illustration using classroom environment data. *Educational Psychology*, 28(5), 583–595.

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis : an introduction to basic and advanced multilevel modeling* (2nd edition ed.). Los Angeles: SAGE.

7 Abbildungsverzeichnis

8 Anhang