



**Universität
Zürich^{UZH}**

Umgang mit genesteten Daten mittels Multilevel Analyse in R

Masterarbeit von

Noah Bosshart

Betreut durch

Prof. Dr. Carolin Strobl

21. Oktober 2019

Inhaltsverzeichnis

1	Abstract	3
2	Einleitung	4
2.1	Genestete Datenstrukturen	4
2.2	Problematik von linearen Modellen	4
3	Multilevel Modelle als Antwort	5
3.1	Wann werden Multilevel Modelle eingesetzt?	5
3.2	Aufbau von Multilevel Modellen	5
3.3	Vergleich von Multilevel Modellen	5
3.4	Kennwerte von Multilevel Modellen	5
4	Programmierung einer Shiny App zur Erläuterung von MLM	5
5	Diskussion	5
6	Literaturverzeichnis	6
7	Abbildungsverzeichnis	6
8	Anhang	6

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 Längsschnittdaten. In diesen Datenstrukturen bestehen gewisse 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 können 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 Analyse 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 Modelle als Antwort

3.1 Wann werden Multilevel Modelle eingesetzt?

3.2 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. (Snijders & Bosker, 2012) (Weitere Guides / Tutorials zu MLM finden)

3.3 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)

3.4 Kennwerte von Multilevel Modellen

4 Programmierung einer Shiny App zur Erläuterung von MLM

5 Diskussion

6 Literaturverzeichnis

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 [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* (2nd edition Aufl.). Los Angeles: SAGE.

7 Abbildungsverzeichnis

8 Anhang