Einführung in Mehrebenenmodelle in R

Julius Fenn [1]

[1]Universität Hamburg
Wissenschaftlicher Mitarbeiter im Institut für Erziehungswissenschaft (EW1)

29.08.2019

Vorstellungsrunde

- Forschungsinteresse
- ggf. Vorerfahrungen mit Mehrebenenanalysen
- ggf. Vorerfahrungen mit R

Aufbau Folien

Verweis auf das zusätzliche Übersichtsdokument [oben rechts]:

→ Zusatz: z.B. **Anhang A**

→ Übungen: **Übung 1: Lineare Regression**

technische Folie, nur Zusatz: → Tech

Hervorhebungen, wie Schwellen für Analyseverfahren in rot [im Text]

Verweis auf den Anhang, Literaturverweise [am Ende]:

Verweise auf den Anhang: ▸ Colemansche Badewanne

→ ohne Seitenangabe: Hox (2010)

→ mit Seitenangabe: Hox (2010), 10f.

→ mehrere Literaturangaben: Raudenbush (2004), 122; Fiege (2016), 58ff.

Abkürzungsverzeichnis

Hinleitende Ausführungen:

„CIPO“-Modell	„Context Input Process Output“-Modell
LM	Lineares Modell

Hierarchisches lineares Modell:

HLM	Hierarchisches lineares Modell
FE	fixen Effekte
ZE	zufällige Effekte

Testung:

DETECT	*Dimensionality Evaluation to Enumerate Contributing Traits*
DIF	*Differential Item Functioning*

Komplexität von Bildungsprozessen I

- Bildungsprozesse werden explizit über Akteur/innen von Bildungseinrichtungen vermittelt und implizit über die unmittelbare soziale Umwelt
  - → zur Analyse von Bildungsprozessen ist es notwendig unterschiedliche analytische Bezugseinheiten auf Makro-, Meso- und Mikroebene zu beachten, die bei der Transformation von Inputs in Outputs eine Rolle spielen und miteinander interagieren

Willems und Budde (2009); Tippelt und Schmidt-Hertha (2018), 375ff.

Komplexität von Bildungsprozessen II: Beispiel

Statistisch → Trennung von Einflussbereichen, die auf den Bildungsprozess wirken, z.B. Modell der Schülerleistung:

$$\mathrm{SL}_{ij} = S_{ij} + \underbrace{C_j + P_j}_{\text{Schulaspekte}} + \varepsilon$$

- SL_{ij}: Schülerleistung eines Schülers i in Schule j
- S_i: Schülercharakteristiken
- C_j: Schulkontext
- P_j: Schulpraxis

zusätzliche werden (implizit) Annahmen an das Modell gestellt:

$$\mathrm{KL}_i = \overline{S_j} + C_j + P_j + \varepsilon$$

Raudenbush (2004), 122; Fiege (2016), 58f.

„CIPO“-Modell I

Anhang A

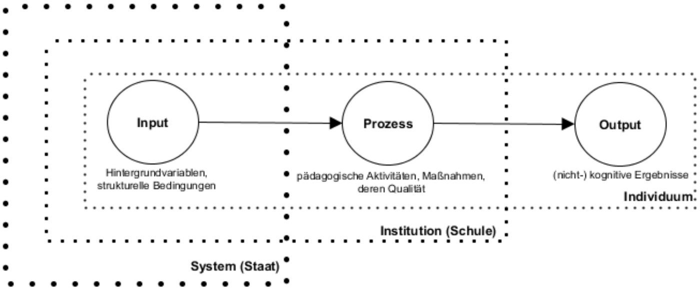
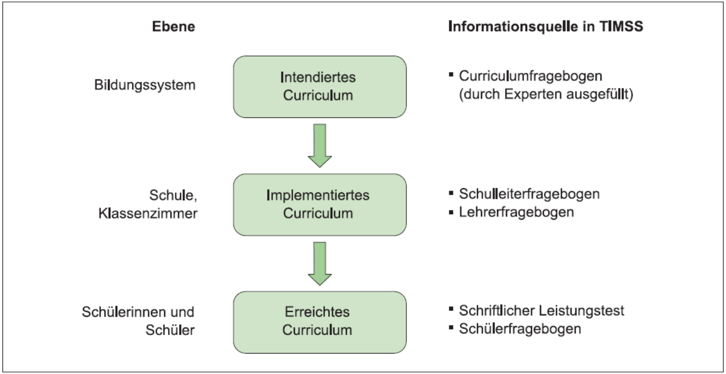


Abbildung: angepasstes „CIPO“-Modell

Anpassung nach: Fischer et al. (2011), 18ff.; Klieme und Vieluf (2013); Tippelt und Schmidt-Hertha (2018), 380ff.

Keller (2014), 43ff.

„CIPO“-Modell II: Anwendung anhand des TIMSS Curriculum-Modells



IEA: Trends in International Mathematics and Science Study © TIMSS 2015

Abbildung: Das TIMSS-Curriculum-Modell

Wendt et al. (2016), 37f.

„CIPO"-Modell III: mögliche Komponenten

Anhang A

Komponenten	**Schüler**	**Schule**	**Staat**
Input	Alter, Geschlecht, Migrationshintergrund; Vorwissen; Einstellungen	soziales Einzuggebiet; Ressourcen; Schulform	Demographie, gesellschaftliche Debatten; Investitionen
Prozess	Lerntempo, Motivation	organisatorische Gestaltung; Elternkooperation; Lehrerkooperation	Regulierungsinstrumente, wie Maß der Schulautonomie; Standardsetzung
Output	Lernergebnisse	Anzahl Abbrecher auf Schulebene; durchschnittliche Schulleistung; Schulklima	Bildungsstand, Wirtschaftswachstum, Wohlstand der Gesellschaft

Tabelle: Einzelne Komponenten des „CIPO"-Modells in Bezug zu den Akteursebenen

Variablen entnommen aus: Klieme und Vieluf (2013), 234; Tippelt und Schmidt-Hertha (2018), 383

„CIPO"-Modell IV: Bezug zu linearen Modellen

Anhang A

Mögliche Variablen, die das Modell der Schulleistung (SL) enthält ergeben sich aus dem „CIPO"-Modell:

$$\mathrm{SL}_{jt} = \overline{S_{ijt}} + \underbrace{C_{jt} + P_{jt}}_{\text{Schulaspekte}} + \varepsilon_{jt}$$

→ hierbei repräsentiert der Index t die Möglichkeit die Schulleistung über die Zeit t zu messen; *Grundidee eines vereinfachten Modells z.B. Coleman-Report - Coleman (1966)*

Generell die Schülerleistung als kumulative Funktion darzustellen basiert auf der Idee, dass sich ein Bildungsprozess additiv (linear) aus verschiedenen Inputs zusammensetzt
→ **Bildungsproduktionsfunktion** (*education production function*)

Todd und Wolpin (2003); Koedel, Mihaly und Rockoff (2015), 181; Anforderungen an die Datengrundlage: Baethge et al. (2010)

Ziel des HLM: Gedankenexperiment

Wie ist es möglich, trotz der Vielzahl an Faktoren, die die Bildungsprozesse beeinflussen, kausale Schlüsse ziehen?

$$Y_Q(A) - Y_Q(B) \qquad , wenn > 0 \text{ Schule A besser}$$

- Y: Testleistung
- Q: Schüler
- A, B: zwei Schulen
- z.B. $Y_Q(A)$: Testleistung Y eines Schülers Q in der Schule A

$\rightarrow$ aufgrund eines fehlenden Datenproblems ist es unmöglich kausale Schlüsse zu ziehen

Rubin, Stuart und Zanutto (2004); Bildungsberichterstattung (2018), 197

statistische Aspekte: Ziele von HLM

- Varianz der AV durch Prädiktoren auf beliebig vielen Ebenen aufklären
- Erforschung von kausaler Heterogenität (*causal heterogeneity*) durch die Spezifizierung durch ebenenübergreifende Interaktionen (*cross level interactions*)
- Bildung von komplexen Bildungsindikatoren mittels zufälliger Effekte (siehe Folie 47)

→ Nichteinbezug von Variablen auf allen Ebenen führt zu einer Verzerrung der geschätzten Koeffizienten (*omitted variable bias*); größte Verzerrung durch Auslassung von Variablen auf den unteren Ebenen (z.B. Schülervariablen) - Kim und Frees (2006)

Steenbergen und Jones (2002); Western (1998)

Colemansche Badewanne

statistische Aspekte: Nichtbeachtung Mehrebenenstruktur

„[T]he more individuals share common experiences due to closeness in space and/or time, the more they are similar, or to a certain extent, duplications of each other“ - Kreft und De Leeuw (1998), 9

→ z.B. in Berlin ist der überwiegende Teil der sozialen Segregation der Grundschulen auf die soziale Segregation der Stadt zurückzuführen - Helbig und Nikolai (2017)

- keine Unabhängigkeit zwischen den Schüler/innen

→ Standardfehler werden zu niedrig geschätzt
  - statistische Tests fälschlicherweise zu häufig als signifikant angenommen (Erhöhung des Fehler 1. Art)
  - Konfidenzintervalle werden zu klein geschätzt

Steenbergen und Jones (2002); Hox (2010), 4f.

statistische Aspekte: Nichtbeachtung Mehrebenenstruktur II

Ist es sinnvoll Daten auf Ebene der Gruppierungsvariable einfach zu aggregieren?

→ (fast) jede Aggregation von Daten führt zu einem Informationsverlust

- implizit wird angenommen, dass die zusammenzufassenden Daten aus homogenen Gruppen kommen, oftmals sind diese jedoch heterogen
- aggregierte Daten weisen beispielsweise bei lineare Regressionen stärke Zusammenhänge und bessere Modellfits auf
  - mögliche Fehlschlüsse die für eine höhere Ebene aufgrund den aggregierten Daten getroffen werden, heißen *atomistic fallacy* oder *individualistic fallacy*

weiterer Aspekt: *borrowing strength* zur Berechnung von zufälligen Effekten nicht möglich, siehe später Folie 47

Clark und Avery (1976); Hox (2010), 2f.

statistische Aspekte: Nichtbeachtung Mehrebenenstruktur III

3 Möglichkeit Hypothesen zu testen:
1) Konfidenzintervall bilden

$$\beta \pm Z_{\alpha/2} * \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{Standardfehler}}$$

$\rightarrow$ enthält den wahren Parameter mit einer Wahrscheinlichkeit von $1 - \alpha$
$\rightarrow$ wenn 0 enthalten nicht signifikanter Prädiktor / Test

2) Teststatistik überschreitet kritischen Wert
$\rightarrow$ Daumenregel: t-value $\geq |2|$

3) Überschreitungswahrscheinlichkeit / p-Werte
$\rightarrow$ $p < \alpha$ Verwerfung der H_0

Zucchini et al. (2009), 161ff.; Fahrmeir, Heumann et al. (2016), 381ff.; für HLM: Luke (2017)

Grundgleichung HLM

Form des „Generalisierten Linearen Gemischten Modells“:

$$Y = \underbrace{X\beta}_{\text{fixen Effekte}} + \underbrace{U\gamma}_{\text{zufälligen Effekte}} + \epsilon$$

- X, U stellen sogenannte Designmatrizen dar, diese beinhalten alle spezifizierten Variablen und Effekte, sowie die Intercepts
- β, γ repräsentieren Vektoren der geschätzten Koeffizienten

Bauer (2003), 139f.; Fahrmeir, Kneib et al. (2013), 361ff.

Lineares Modell: Grundannahmen

Lineare Zusammenhänge zwischen einer abhängigen metrischen Variablen und einer (oder mehreren) unabhängigen Variablen sollen bestimmt werden, und es lassen sich Werte der abhängigen Variablen prognostizieren

$$\text{UV}_{\text{Unabhängige Variable}} \rightarrow \text{AV}_{\text{Abhängige Variable}}$$

- β-Koeffizienten sind für alle Aggregateinheiten sind gleich
- Linearität: die abhängige Variable und die unabhängige Variable verändern sich nur in konstanten Relationen
- Residuen zwischen den Merkmalsträgern sind voneinander unabhängig

  $\rightarrow$ Annahme $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$ nicht in Mehrebenenstruktur gegeben

Zucchini et al. (2009), 345ff.; Vorlesung Lineare Modelle Skript 2009

Lineares Modell: Einfache lineare Regression

$$Y_i = \beta_0 + \beta 1 x_i + \epsilon_i \qquad i = 1, \ldots, n$$

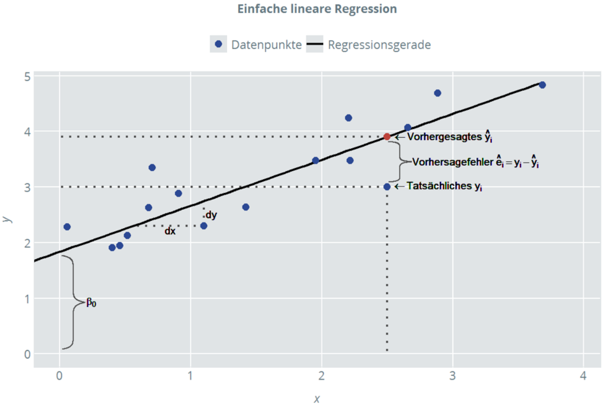
$$Y = \underbrace{X\beta}_{\text{fixen Effekte}} + \text{keine ZE} + \epsilon$$

- Y_i: abhängige Variable, Zielgröße
- x_i: unabhängige Variable, feste bekannte Einflussgröße
- ϵ: Zufallsfehler

- $\beta_0, \beta_1, \sigma^2$: unbekannte Parameter
- n: Anzahl der Beobachtungen

Designmatrix

Lineares Modell: Residuum



Schätzer nach Methode der kleinsten Quadrate:
$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} = \sum_{i=1}^{n}(y_i - \hat{y}_i)$

Lineares Modell: Grundannahmen II

Tech

1. Anzahl der Prädiktoren ist kleiner als Anzahl der Fälle (Identifizierbarkeit)
2. Annahme von Linearität
3. Erwartungswert der Fehler ist 0: $E(\epsilon_i) = 0$
4. Varianzen der Fehler sind gleichverteilt (Homoskedastizität): $Var(\epsilon_i) = \sigma^2$
5. Keine Korrelation zwischen den Fehlern: $Cov(\epsilon_i, \epsilon_j) = 0$, für ij
6. Keine exakte Multikollinearität der Prädiktoren
7. Normalverteilung der Fehler ϵ_i

$\rightarrow$ treffen Annahmen 2 bis 6 zu ist das Modell BLUE (kleinste Varianz, linear, nicht verzerrt)

▸ weitere Schlussfolgerungen

Einschub: Betrachtung Datensatz

Übersicht Datensatz (1. K.)

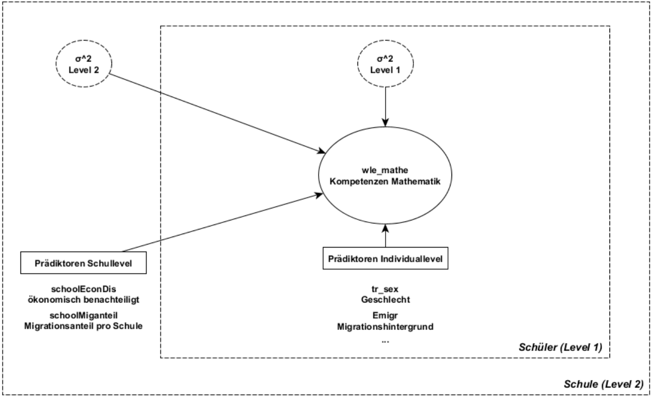


Abbildung: Konzeptuelles Modell der zu analysierenden Mehrebenenstruktur des Workshops. Idee folgt Schmidt, Zlatkin-Troitschanskaia und J.-P. Fox 2016, 8

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Lineares Modell: Übung 1 (Teil 1)

Übung 1 (2. K.)

Aufgabe: Stellen Sie eine lineare Regression mit der Variable auf, die die Anzahl der Bücher zu Hause angibt. Verwenden Sie dafür die betreffende Vorlage in der R-Datei `02_Uebung_1`. Vergleichen Sie die Ergebnisse mit dem zusätzlichen Dokument zweites Kapitel.

Lineares Modell: Übung 1 (Teil 1)

Übung 1 (2. K.)

$$Y_i = \beta_0 + \beta_1 x_{25\ \text{Bücher}} + \beta_2 x_{100\ \text{Bücher}} + \beta_3 x_{200\ \text{Bücher}} + \beta_4 x_{>200\ \text{Bücher}} + \epsilon_i$$

- x_{SBuecher}: unabhängige Variable $\rightarrow$ **Dummy-Kodierung**

kategoriale Regressoren mit k Kategorien werden durch einen Vektor von $m = k - 1$ Dummy-Variablen kodiert:

$$X^{(i)} = \begin{cases} 0, & \text{falls Kateogrie i beobachtet wird} \\ 1, & \text{sonst} \end{cases}$$

k-te Kategorie entspricht der Referenzkategorie (Dummy-Variablen haben den Wert 0)

Fahrmeir, Heumann et al. (2016), 452f.

Lineares Modell: Übung 1 (Teil 2)

Übung 1 (2. K.)

Aufgabe: Stellen Sie eine lineare Regression mit der Variable auf, die den höchsten ökonomischen Status angibt. Zentrieren Sie diese Variable um ihren Mittelwert. Verwenden Sie dafür die betreffende Vorlage in der R-Datei `02_Uebung_1`. Vergleichen Sie die Ergebnisse mit dem zusätzlichen Dokument zweites Kapitel.

Lineares Modell: Übung 1 (Teil 2)

Übung 1 (2. K.)

$$Y_i = \beta_0 + \beta_1 x_{\text{EHiseiGM}} + \epsilon_i$$

- x_{EHiseiGM}: unabhängige Variable → **Mittelwertzentrierung** (grandmean centering)
- Schüler/in erhält den Wert 0, wenn die Variable genau der durchschnittlichen Variable der gesamten Stichprobe entspricht
  - → negative Werte: niedrigeren EHisei als der mittlere EHisei der gesamten Stichprobe
  - → positive Werte: höheren EHisei als der mittlere EHisei der gesamten Stichprobe

Zentrierung wird nochmals auf Folie 28 weiter ausgeführt

Fahrmeir, Heumann et al. (2016), 452f.

Lineares Modell: Übung 1 (Teil 3)

Übung 1 (2. K.)

Aufgabe: Stellen Sie eine lineare Regression mit den aus Teil 1 und Teil 2 verwendeten Variablen auf. Spezifizieren Sie eine Interaktion für beide Variablen. Verwenden Sie dafür die betreffende Vorlage in der R-Datei `02_Uebung_1`. Vergleichen Sie die Ergebnisse mit dem zusätzlichen Dokument zweites Kapitel.

Lineares Modell: Übung 1 (Teil 3)

Übung 1 (2. K.)

$$Y_i = \beta_0 + \beta_1 x_{\text{EHiseiGM}} + \beta_2 x_{\text{25 Bücher}} + \beta_3 x_{\text{100 Bücher}} + \beta_4 x_{\text{200 Bücher}} + \beta_5 x_{>\text{200 Bücher}} + \underbrace{\gamma(x_{\text{EHiseiGM}} * D)}_{\text{Interaktion}} + \epsilon_i$$

- $\gamma(x_{\text{EHiseiGM}} * D)$: Interaktion $\rightarrow$ Zusammenführung einer kontinuierlichen und diskreten Variable, die Interaktion erlaubt eine Veränderung in dem Steigungskoeffizienten γ
  - auch genannt *interaction variable, slope-indicator variable*

Beispiel: Linear Regression Using Dummy Variables 2017

Zentrierung

- unterschiedliche Zentrierung (centering) auf L1 und L2 möglich und sinnvoll
- dienen der Festlegung eines inhaltlich plausiblen Nullpunkts für die unabhängige(n) Variable(n)

zwei Möglichkeiten:

1. Zentrierung am Gesamtmittelwert (grandmean centering)
2. Zentrierung am gruppenspezifischen Mittelwert der Variablen (groupmean centering)

How to center in multilevel models 2018; Bauer (2003), 139f.; Fahrmeir, Kneib et al. (2013), 361ff.

Zentrierung am Gesamtmittelwert (grandmean centering, GMC)

- sowohl auf Individual- als auch auf Kontextebene möglich
- Gesamtdurchschnittswert einer UV wird von den individuellen Beobachtungswerten abgezogen
- Interpretation siehe Folie 25

Bauer (2003), 139f.; Fahrmeir, Kneib et al. (2013), 361ff.

Zentrierung am gruppenspezifischen Mittelwert der Variablen (centering within cluster, CWC)

- nur auf Individualebene möglich
- Durchschnittswert einer UV einer Schule wird von den individuellen Beobachtungswerten abgezogen
- Schüler/in erhält den Wert 0, wenn die Leistung genau der durchschnittlichen Leistung seiner Schule entspricht
  - → negative Werte: schlechtere Leistungen als die mittlere Leistung der Schule
  - → positive Werte: bessere Leistungen als die mittlere Leistung der Schule

Bauer (2003), 139f.; Fahrmeir, Kneib et al. (2013), 361ff.

Schulleistungsuntersuchungen (*large-scale-assessments*, LSA)

Tech

Schulleistungsuntersuchungen weisen ein besonderes Design auf, dass den Einbezug von (1) Gewichten und (2) einer passenden Varianzschätzung benötigt

sichtbar in der Schätzformel des Standardfehlers:

$$\mathrm{VAR}_{\mathrm{tot}} = \underbrace{\mathrm{VAR}_{\mathrm{sml}}[\theta]}_{\text{sampling error}} + \underbrace{(1 + \frac{1}{M}) * \mathrm{VAR}_{\mathrm{imp}}[\theta]}_{\text{imputation error}}$$

Caro und Biecek (2017); L. Rutkowski, Davier und D. Rutkowski (2013), 389ff.

Gewichte → Sampling Error

- korrigiert disproportionale Ziehungswahrscheinlichkeiten
- in IEA-Studien HOUWGT, TOTWGT und SENWGT

→ das einfachste Gewicht ist das Gewicht einer einzelnen Schüler/innen s_i (wobei p die Ziehungswahrscheinlichkeit angibt):

$$s_i = \frac{1}{p}$$

ausgehend davon werden komplexere Gewichte berechnet und möglich Ausfallprozesse berücksichtigt

→ selbst wurde ein „normalisiertes" Gewicht (*student house weight*) für die Analysen berechnet: $N * \frac{x_{\mathrm{wgtSTUD}}}{\sum x_{\mathrm{wgtSTUD}}}$

Martin, Mullis und Hopper (2016), 3.1ff.; Leeuw, Hox und Dillman (2012), 317ff.

Plausible Values → Imputation Error

- fünf Plausible Values (PVs), die Auskunft über die Leistung eines Schülers geben, wobei ein Schüler in der Stichprobe Personen mit ähnlichen Eigenschaften in der Population repräsentiert (unter Berücksichtigung von Hintergrundmerkmalen)
- Stichprobe von Testaufgaben und Stichprobe von Schülern, obwohl Aussagen über Populationen gemacht werden sollen

→ durch Berücksichtigung von PVs nähert man sich dieser wahren Varianz an, korrekte Berechnung von Standardfehlern

Aufbau

Modellansatz wird auch als *multilevel linear model*, *random-coefficient regression model*, *mixed model* oder *variance component model* bezeichnet

folgende Modelle, werden neben weiteren theoretischen Aspekten ausgeführt:

- Nullmodell
- Random Coefficient Model
- corss level interaction
- Ausblick Erweiterungen (wenn Zeit)

Voraussetzungen HLM

- Kriterium (AV) liegt auf der untersten Ebene vor und muss numerisch sein
- Prädiktoren (UVs) können auf allen Ebenen vorliegen

→ Datensatz muss eine (relevante) Mehrebenenstruktur aufweisen - siehe Ausführungen auf Folie 44

Modellierungsstrategien

„Statistical modeling is a process that is guided by substantive theory, the results of exploratory analysis, and results from fitting various models to the data.“

- *step-up method*: ein einfaches Level-1 Modell wird spezifiziert, welches einen variierenden Intercept für die Gruppierungsvariable enthält, wobei weitere Prädiktoren / Effekte schrittweise hinzugefügt werden
- *top-up approach*: das komplexeste Modell wird aufgestellt, wobei schrittweise die Komplexität reduziert wird
  - dieser Ansatz eignet sich besser für longitudinale Daten (falls nicht-lineare Zusammenhänge)

L. Rutkowski, Davier und D. Rutkowski (2013), 408f.

Modellierungsstrategien II

von den Autoren Kim, Anderson, Keller (2013) wird folgender Ablauf vorgeschlagen:

1. explorative Datenanalyse
2. Berechnung eines Nullmodells
3. Einbezug von fixen Effekten (Prädiktoren)
4. Ausbau der zufälligen Effektstruktur (Prädiktoren wirken unterschiedlichen innerhalb Gruppierungsvariable)
5. Testung der spezifizierten fixen Effekte
6. Testung der spezifizierten zufälligen Effekte
7. Diagnose des finalen Modells durchführen

L. Rutkowski, Davier und D. Rutkowski (2013), 409

Modellierungsstrategien III

Ziel des Einbezugs von weiteren Prädiktoren, sowie einer Struktur der zufälligen Effekte siehe Folie 12

weiterhin kann über testtheoretische Gütemaße, wie der Konstruktvalidität argumentiert werden:

- Konstruktvalidität basiert auf „integration of any evidence that bears on the interpretation or meaning of the test scores" - Messick (1994), 3
    - Einbezug weiterer Variablen, deren Einfluss auf die AV theoretisch prognostiziert wird, erlaubt eine Stärkung der Konstruktvalidität (der skalierten Variable)

Moosbrugger und Kelava (2012), 143ff.

R-Paket lme4: Einführung

Tech

lme4: Linear mixed-effects models
→ Paket ermöglicht den Einbezug von sogenannten fixen und zufälligen Effekten, in R können Informationen zu dem Paket durch „?lme4“ abgerufen werden

Parameterschätzung: Parameter werden mittels einer *maximum likelihood* Schätzung berechnet → diese bezieht zur Berücksichtigung der zufälligen Effekte einen sogenannten *penality term* ein
→ standardmäßig verwendet das Paket einen REML-Schätzer - Steenbergen und Jones (2002), 225f.

Maximierung der *(penalized) maximum likelihood* Schätzung mittels dem BOBYQA Algorithmus

Bates et al. (2014)

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

R-Paket lme4: Syntax

Formula	Alternative	Meaning
`(1|g)`	`1+(1|g)`	Random intercept with fixed mean
`0+offset(o)+(1|g)`	`-1+offset(o)+(1|g)`	Random intercept with *a priori* means
`(1|g1/g2)`	`(1|g1)+(1|g1:g2)`	Intercept varying among g1 and g2 within g1
`(1|g1)+(1|g2)`	`1+(1|g1)+(1|g2)`	Intercept varying among g1 and g2
`x+(x|g)`	`1+x+(1+x|g)`	Correlated random intercept and slope
`x+(x||g)`	`1+x+(1|g)+(0+x|g)`	Uncorrelated random intercept and slope

Abbildung: Beispiel der Modellformeln der HLM

Bates et al. (2014), 6

Nullmodell

- überprüft wird, wie sich die Varianz der abhängigen Variablen (hier: Kompetenzen Mathematik) auf den beiden Ebenen verteilt
- hierfür: Modellierung des Mittelwerts der abhängigen Variablen auf beiden Ebenen
- daraus kann abgeleitet werden, ob es „überhaupt sinnvoll“ ist ein Mehrebenenmodell zu berechnen

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \epsilon_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

Steenbergen und Jones (2002), 224; Hox (2010), 14f.

Nullmodell II

Zusammenfassung des Level 1 und Level 2 Modells:

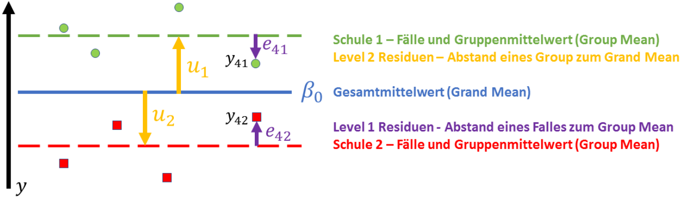
$$Y_{ij} = \gamma_{00} + \underbrace{u_{0j} + \epsilon_{ij}}_{\text{Fehlerterme}}$$

- Y_{ij}: abhängige Variable (hier: Kompetenzen Mathematik)
- γ_{00}: Gesamtmittelwert $\rightarrow$ fixer Effekt
- u_{0j}: Variation auf Schulebene im Level 1 Intercept
- ϵ: Level 1 Fehlerterm

- Index i: Individuum (Schüler/in)
- Index j: Gruppe (Schule)

Nullmodell III: graphische Darstellung

$$y_{ij} = \beta_0 + u_j + e_{ij}$$



$$Punkte_{ij} = Mittelwert\ aller\ Schüler + Fehler_{L2} + Fehler_{L1}$$

Abbildung: Multilevel Nullmodell: Beispielgrafik mittels zwei Schulen

Level 1 Fehler: $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
Level 2 Fehler: $u_j \sim N(0, \sigma_u^2)$

intra-class correlation coefficient (ICC)

der ICC gibt den Anteil der Varianz an, welche durch die Gruppierungsvariable u erklärt wird (σ_u^2 gibt die Variabilität zwischen den Gruppen / Clustern an):

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

$$ICC = \frac{\text{Varianz}_{L2}}{\text{Varianz}_{L2} + \text{Varianz}_{L1}}$$

- Hox (2010): Abhängigkeit zwischen Beobachtungen innerhalb eines Clusters drückt sich in einem nicht-trivalen ICC aus
- Geldhof et al (2014): alle ICC die nicht trivial und beispielsweise größer als .05 sind müssen berücksichtigt werden

Geldhof, Preacher und Zyphur (2014); Hox (2010), 4, 15, 33ff.

Nullmodell: Übung 2

Übung 2 (3. K.)

Aufgabe: Stellen Sie das sogenannte Nullmodell auf. Verwenden Sie dafür die betreffende Vorlage in der R-Datei `03_Uebung_2`.
Vergleichen Sie die Ergebnisse mit dem zusätzlichen Dokument drittes Kapitel.
Wie schätzen Sie die Höhe des ICC ein? Berechnen Sie den ICC per Hand.

intra-class correlation coefficient (ICC) II

Tech

Für Mehrebenenanalysen gibt es keine einzelne Zahl, die das Ausmaß der Variabilität zwischen den Gruppen genau angeben kann. So beschreibt Goldstein et al. (2002), dass in „multilevel modelling, the residual variation in a response variable is split into component parts that are attributed to various levels. [..] Such a measure however only makes sense in simple variance components" - Goldstein, Browne und Rasbash (2002), 223

weiterführende Diskussion findet sich unter: Intraclass Correlation Coefficient in mixed model with random slopes 2015

Berechnung des *Conditional ICC* - Nakagawa, Johnson und Schielzeth (2017):

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_f^2 + \sigma_e^2} \quad , wobei \ \sigma_f^2 = Var(\sum_h^k \beta_h x_{h_{ij}})$$

Geldhof, Preacher und Zyphur (2014); Hox (2010), 4, 15, 33ff.

zufällige Effekte: Berechnung

Anhang B, Tech

$$\text{BLUP}_k = \text{BLUE}_k * \text{shrinkage factor}_k = \hat{\mu} + (\hat{\mu}_k - \hat{\mu}) * \left(\frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \frac{\hat{\sigma}_E^2}{r_k}} \right)$$

Dabei ist der *shrinkage factor* in folgenden Fällen groß (nahe 1) $\rightarrow$ der Populationsmittelwert ist kein guter Indikator für den *Best Linear Unbiased Predictor*:

- die geschätzten Gruppenvarianzen $\hat{\sigma}_G^2$ ist groß
- der Fehleranteil in der Gruppenvarianz $\hat{\sigma}_E^2$ ist klein
- die Gruppengrößen der Cluster k ist groß (r_k)

Galwey (2014), 168f.; Hox (2010), 30f.; Beispiel: Sanders und Horn (1994), 307f.

zufällige Effekte: Anwendung Hinführung

zufällige Effekte: „a way to combine information from different levels within a grouping variable“ - G. Fox, Negrete-Yankelevich und Sosa (2015), 312

dabei können sich zufällige Effekte als Werte vorgestellt werden, die zwischen zwei möglichen Extremen liegen:

- Gruppierungsvariablen werden völlig ignoriert (*fully pooled estimation*): $y_{ij} = \beta_0 + \epsilon_{ij}$
- für jede Gruppe (Cluster) i eine eigene Regression berechnen (*fully unpooled estimation*): $y_{ij} = \beta_i + \epsilon_{ij}$

Fahrmeir, Kneib et al. (2013), 355ff.; G. Fox, Negrete-Yankelevich und Sosa (2015), 312f.

zufällige Effekte: Anwendung

Darstellung der einzelnen Schulen mit Fehlerbalken (siehe Übersichtsskript Seite 7) ist die einzige Möglichkeit zu entscheiden, ob sich die zufälligen Effekte signifikant von null unterscheiden → Vorgehen entspricht der klassischen Bildung eines Konfidenzintervalls (siehe Folie 15) - G. Fox, Negrete-Yankelevich und Sosa (2015), 312ff.

! für das R-Paket lme4 standardmäßig kein Signifikanzniveau für geschätzten Parameter angegeben → aufgrund fehlender Daten, unbalancierte Gruppen und durch die Spezifizierung zufälliger Effekte ist es nicht klar, was der Freiheitsgrad eines HLM ist - siehe ausführlicher ▸ Quellen Begründung

Random Coefficient Model

- Aufnahme eines Prädiktors auf L1, der (theoretisch) die Varianz der abhängigen Variablen auf L1 erklären kann

→ es unterscheiden sich nur die Intercepts (Mittelwerte) der abhängigen Variablen zwischen den Schulen

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_1 x_{\text{EHiseiGM}} + \epsilon_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\text{Level 2: } \beta_{1j} = \gamma_{10}$$

Steenbergen und Jones (2002), 224

Random Coefficient Model II

Zusammenfassung des Level 1 und Level 2s Modells:

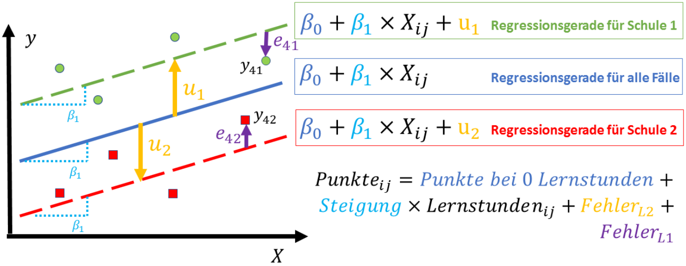
$$Y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{10}\, x_{\text{EHiseiGM}} + \epsilon_{ij}$$

- Y_{ij}: abhängige Variable (hier: Kompetenzen Mathematik)
- γ_{00}: Gesamtmittelwert $\rightarrow$ fixer Effekt
- u_{0j}: Variation auf Schulebene im Level 1 Intercept
- $\gamma_{10}\, x_{\text{EHiseiGM}}$: Steigungskoeffizient $\rightarrow$ fixer Effekt
- ϵ: Level 1 Fehlerterm

- Index i: Individuum (Schüler/in)
- Index j: Gruppe (Schule)

Random Coefficient Model III: graphische Darstellung

$$y_{ij} = \beta_0 + \beta_1 \times X_{ij} + u_j + e_{ij}$$



$$Punkte_{ij} = Punkte\ bei\ 0\ Lernstunden + Steigung \times Lernstunden_{ij} + Fehler_{L2} + Fehler_{L1}$$

Abbildung: Multilevel *Random Coefficient Model*: Beispielgrafik mittels zwei Schulen

Level 1 Fehler: $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
Level 2 Fehler: $u_j \sim N(0, \sigma_u^2)$

Random Coefficient Model: Übung 3

Übung 3 (4. K.)

Aufgabe: Stellen Sie das sogenannte *Random Coefficient Model* mit folgenden eben berechneten Variablen auf:
`EHisei.gmc, EHisei.cm, EHisei.cwc`
Verwenden Sie dafür die betreffende Vorlage in der R-Datei `04_Uebung_3`. Vergleichen Sie die Ergebnisse mit dem zusätzlichen Dokument viertes Kapitel.

Random Coefficient Model: random slopes

- Aufnahme eines Prädiktors auf L1, der (theoretisch) die Varianz der abhängigen Variablen auf L1 erklären kann

→ es unterscheiden sich die Intercepts (Mittelwerte) und die Steigungskoeffizienten der abhängigen Variablen zwischen den Schulen

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} x_{\text{EHiseiGM}} + \epsilon_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\text{Level 2: } \beta_{1j} = \gamma_{10} + u_{1j}$$

`Steenbergen und Jones (2002), 221f.`

Random Coefficient Model: random slopes II

Zusammenfassung des Level 1 und Level 2s Modells:

$$Y_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})\, x_{\text{EHiseiGM}} + \epsilon_{ij}$$

- Y_{ij}: abhängige Variable (hier: Kompetenzen Mathematik)
- γ_{00}: Gesamtmittelwert $\rightarrow$ fixer Effekt
- u_{0j}: Variation auf Schulebene im Level 1 Intercept
- $\gamma_{10}\, x_{\text{EHiseiGM}}$: Steigungskoeffizient $\rightarrow$ fixer Effekt
- $u_{1j}\, x_{\text{EHiseiGM}}$: Steigungskoeffizient für Schule j $\rightarrow$ zufälliger Effekt
- ϵ: Level 1 Fehlerterm

- Index i: Individuum (Schüler/in)
- Index j: Gruppe (Schule)

Random Coefficient Model: random slopes III: graphische Darstellung

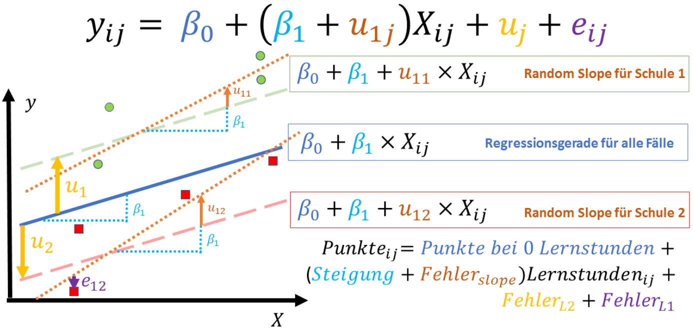


Abbildung: Multilevel *Random Coefficient Model* - Spezifizierung eines random slopes: Beispielgrafik mittels zwei Schulen

Level 1 Fehler: $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

Level 2 Fehler: $u_0j \sim N(0, \tau_{00})$, $u_1j \sim N(0, \tau_{11})$, $Cov(u_0j, u_1j) = \tau_{01}$

Random Coefficient Model random slopes: Übung 4

Übung 4 (5. K.)

Aufgabe: Stellen Sie das sogenannte *Random Coefficient Model* mit zufälligen Intercepts und Steigungskoeffizienten auf. Sie können sich die Syntax der lme4 Pakets auf Folie 40 zu Hilfe nehmen. Versuchen Sie weiterhin ein Modell nur mit zufälligen Intercepts und ein Modell nur mit zufälligen Steigungskoeffizienten zu spezifizieren.
Verwenden Sie dafür die betreffende Vorlage in der R-Datei `05_Uebung_4`. Vergleichen Sie die Ergebnisse mit dem zusätzlichen Dokument viertes Kapitel.

Random Coefficient Model: random slopes VI: zufällige Effekte

das HLM nimmt an, dass die zufälligen Effekte u_{0j} und u_{1j} aus einer bivariaten Normalverteilung kommen:

$$u_j = \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .15 & -.03\,(\text{Kor}) \\ -.0001 & .0001 \end{pmatrix} \right)$$

→ hierbei gibt der Wert vor „(Kor)“ die Korrelation nach Pearson an (erechnet durch $\rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$)

Fahrmeir, Kneib et al. (2013), 358

Modellselektion Verfahren

trotz Probleme der Berechnungen von klassischen p-Werte (siehe Folie 49), bieten sich einzelne Verfahren an, um HLM miteinander zu vergleichen:

- Informationskriterien, wie *Akaike information Criterion* (AIC) und *Bayesian Information Criterion* (BIC) → kleinere Werte drücken bessere Modellanpassung aus - Fahrmeir, Kneib et al. (2013), 664ff.
- Verwendung eines Likelihood-Ratio-Tests, hierbei berechnet sich der χ^2 Wert aus $-2 * log(L(\hat{\theta}_s)) - log(L(\hat{\theta}_g)) \sim \chi^2$; Voraussetzung zum Vergleich eines einfacheren Modells s mit einem generellen Modell g ist, dass beide verschachtelt sind - Fahrmeir, Kneib et al. (2013), 662f.
- es ist möglich für die einzelnen Koeffizienten Konfidenzintervalle zu berechnen / zu simulieren - siehe im Detail „?confint.merMod“, sowie Baayen, Davidson und Bates (2008)

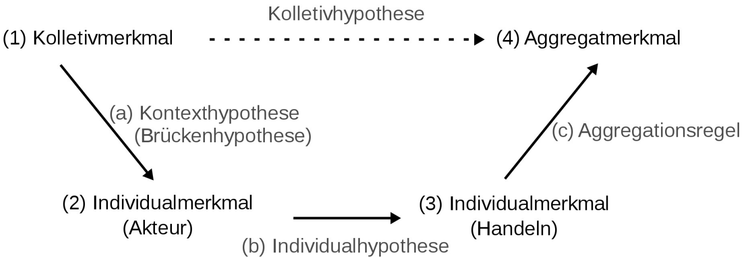
Erweiterungen des HLM

- das HLM ist ein Teilmodell der generalisierten linearen Modelle (GLM), damit kann es auch mit binären AV umgehen (Stichwort logistische Regression) - z.B. Fahrmeir, Kneib et al. (2013)
- mit dem R-Paket MplusAutomation lassen sich Analysen leicht von R in Mplus übertragen
- es ist möglich äquivalente latente Wachstumskurvenmodelle für HLM zu spezifizieren - z.B. Bauer (2003)
- innerhalb der *item response theory* werden GLM verwendet und Personenfähigkeiten und Itemschwierigkeiten zu berechnen - z.B. De Boeck und Wilson (2004)
- ...

Literatur

- Baayen, Harald, Douglas Davidson und Douglas Bates (2008). „Mixed-effects modeling with crossed random effects for subjects and items". In: *Journal of memory and language* 59.4, S. 390–412.
- Baethge, Martin et al. (2010). *Indikatorenentwicklung für den nationalen Bildungsbericht "Bildung in Deutschland"*. BMBF.
- Bates, Douglas et al. (2014). „Fitting linear mixed-effects models using lme4". In: *Journal of Statistical Software.*
- Bauer, Daniel (2003). „Estimating multilevel linear models as structural equation models". In: *Journal of Educational and Behavioral Statistics* 28.2, S. 135–167.
- Bildungsberichterstattung, Autorengruppe (2018). *Bildung in Deutschland 2018: ein indikatorengestützter Bericht mit einer Analyse zu Wirkung und Erträgen von Bildung.* https://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2018/bildung-in-deutschland-2018. Accessed: 2018-08-22.
- Caro, Daniel und Przemyslaw Biecek (2017). „intsvy: An R package for analysing International large-scale assessment data". In: *Journal of Statistical Software* 81.1, S. 1–44.
- Clark, William und Karen Avery (1976). „The effects of data aggregation in statistical analysis". In: *Geographical Analysis* 8.4, S. 428–438.
- Coleman, James (1966). „Equality of educational opportunity". In:
- — (1991). *Grundlagen der Sozialtheorie. Band 1 Handlungen und Handlungssysteme.* Oldenbourg Verlag München.
- De Boeck, Paul und Mark Wilson (2004). *Explanatory item response models.* Springer.
- Ditton, Hartmut (2013). „Bildungsgerechtigkeit: Analysen zu einer vergessenen Voraussetzung". In: *Dialog der Religionen: eine interdisziplinäre Annäherung.* LIT Verlag Münster, S. 155–185.
- Fahrmeir, Ludwig, Christian Heumann et al. (2016). *Statistik: Der Weg zur Datenanalyse.* Springer-Verlag.
- Fahrmeir, Ludwig, Thomas Kneib et al. (2013). *Regression Models, Methods and Applications.* Springer.
- Fiege, Christiane (2016). „Faire Vergleiche bei Vergleichsarbeiten: Möglichkeiten und Grenzen". In: *Sozialindices für Schulen: Hintergründe, Methoden und Anwendung.* Waxmann Verlag, S. 71–96.
- Fischer, Natalie et al. (2011). *Ganztagsschule: Entwicklung, Qualität, Wirkungen: Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagsschulen (StEG).* Beltz Juventa.
- Fox, Gordon, Simoneta Negrete-Yankelevich und Vinicio Sosa (2015). *Ecological statistics: contemporary theory and application.* Oxford University Press, USA.
- Galwey, Nicholas (2014). *Introduction to Mixed Modelling Beyond Regression and Analysis of Variance.* John Wiley & Sons.
- Geldhof, John, Kristopher Preacher und Michael Zyphur (2014). „Reliability estimation in a multilevel confirmatory factor analysis framework.". In: *Psychological methods* 19.1, S. 72–91.
- Goldstein, Harvey, William Browne und Jon Rasbash (2002). „Partitioning variation in multilevel models". In: *Understanding statistics: statistical issues in psychology, education, and the social sciences* 1.4, S. 223–231.
- Helbig, Marcel und Rita Nikolai (2017). *Alter Wolf im neuen Schafspelz? Die Persistenz sozialer Ungleichheiten im Berliner Schulsystem.* Techn. Ber. WZB Discussion Paper.
- Hox, Joop (2010). *Multilevel analysis: Techniques and applications.* Routledge.
- Keller, Florian (2014). *Strukturelle Faktoren des Bildungserfolgs: Wie das Bildungssystem den Übertritt ins Berufsleben bestimmt.* Springer-Verlag.
- Kim, Jee-Seon und Edward Frees (2006). „Omitted Variables in Multilevel Models". In: *Psychometrika* 4.71, S. 659–690.
- Klieme, Eckhard und Svenja Vieluf (2013). „Schulische Bildung im internationalen Vergleich". In: *PISA 2009-Impulse für die Schul-und Unterrichtsforschung.* Beltz Juventa, S. 229–246.
- Koedel, Cory, Kata Mihaly und Jonah Rockoff (2015). „Value-added modeling: A review". In: *Economics of Education Review* 47, S. 180–195.
- Kreft, Ita und Jan De Leeuw (1998). *Introducing multilevel modeling.* Sage.
- Leeuw, Edith, Joop Hox und Don Dillman (2012). *International handbook of survey methodology.* Routledge.
- Luke, Steven G (2017). „Evaluating significance in linear mixed-effects models in R". In: *Behavior Research Methods* 49.4, S. 1494–1502.
- Martin, Michael, Ina Mullis und Martin Hopper (2016). *Methods and procedures in TIMSS 2015.* Techn. Ber. International Association for the Evaluation of Educational Achievement.
- Messick, Samuel (1994). *Validity of psychological assessment: Validation of inferences from persons responses and performances as scientific inquiry into score meaning.* Techn. Ber. Conference on Contemporary Psychological Assessment.
- Moosbrugger, Helfried und Augustin Kelava (2012). *Testtheorie und Fragebogenkonstruktion.* Springer.
- Nakagawa, Shinichi, Paul CD Johnson und Holger Schielzeth (2017). „The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded". In: *Journal of the Royal Society Interface* 14.134, S. 1–11.
- Raudenbush, Stephen (2004). „What are value-added models estimating and what does this imply for statistical practice?". In: *Journal of Educational and Behavioral Statistics* 29.1, S. 121–129.
- Rubin, Donald B, Elizabeth A Stuart und Elaine L Zanutto (2004). „A potential outcomes view of value-added assessment in education". In: *Journal of educational and behavioral statistics* 29.1, S. 103–116.
- Rutkowski, Leslie, Matthias Davier und David Rutkowski (2013). *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis.* CRC Press.
- Sanders, William und Sandra Horn (1994). „The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment". In: *Journal of Personnel Evaluation in education* 8.3, S. 299–311.
- Schmidt, Susanne, Olga Zlatkin-Troitschanskaia und Jean-Paul Fox (2016). „Pretest-Posttest-Posttest Multilevel IRT Modeling of Competence Growth of Students in Higher Education in Germany". In: *Journal of educational measurement* 53.3, S. 332–351.
- Steenbergen, Marco und Bradford Jones (2002). „Modeling multilevel data structures". In: *american Journal of political Science*, S. 218–237.
- Tippelt, Rudolf und Bernhard Schmidt-Hertha (2018). *Handbuch Bildungsforschung.* Springer.
- Todd, Petra und Kenneth Wolpin (2003). „On the specification and estimation of the production function for cognitive achievement". In: *The Economic Journal* 113.485, F3–F33.
- Wendt, Heike et al. (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich.* Waxmann Verlag.
- Western, Bruce (1998). „Causal heterogeneity in comparative research: A Bayesian hierarchical modelling approach". In: *American Journal of Political Science* 42.4, S. 1233–1259.
- Willems, Katharina und Jürgen Budde (2009). *Bildung als sozialer Prozess.* Beltz Juventa.
- Zucchini, Walter et al. (2009). *Statistik für Bachelor-und Masterstudenten: eine Einführung für Wirtschafts-und Sozialwissenschaftler.* Springer-Verlag.

Colemansche Badewanne: Modell des Makro-Mikro-Makro-Schemas



Coleman (1991), 10ff.

Zurück

weitergehend: Philosophie <-> Bildungsforschung: Ditton (2013)

Designmatrix X

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Zurück

weitere Schlussfolgerungen, die sich aus Modellannahmen ergeben

- Residuen sollten nur zufällige Effekte und keine systematischen Effekte erfassen ($E(\epsilon_i) = 0$)
  - $\rightarrow$ möglichst randomisierte Stichprobe wird angestrebt
- z.B. $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{nS_x^2}$
  - $\rightarrow$ Aufbau von Schätzern betrachten: je größer die Streuung der x-Werte und je größer der Stichprobenumfang desto genauer ist die Schätzung (Theorie: *restriction of range*); *vergleiche Folie Nummer 15*
- ...

Zurück

Begründung keine p-Werte in HLM

Kritik an p-Werten wird von den Programmierern des lme4 Pakets unter folgenden zwei Links im Details ausgeführt:
http://t1p.de/bbolker-FAQ-pvalues
http://t1p.de/bbolker-explanation-pvalues

Quellen:

Baayen, Davidson und Bates (2008)

Luke (2017)

Zurück