



■ Prof. Dr. Stefan König

„Mehrebenen-Analysen:  
Unterrichtsforschung  
im Fach Sport zur Effektivität  
von Trainingsprozessen“

■ Vortrag an der PH Weingarten am  
06.12.2017



■ Prof. Dr. Stefan König

„Mehrebenen-Analysen:  
Unterrichtsforschung  
im Fach Sport zur Effektivität  
von Trainingsprozessen“

■ Vortrag an der PH Weingarten am  
06.12.2017

# Agenda

Warum MLM?

Das Basismodell

Erweiterungen des Basismodells

Ein paar allgemeine ‚Musts‘ von MLM

Zusammenfassung

# Agenda

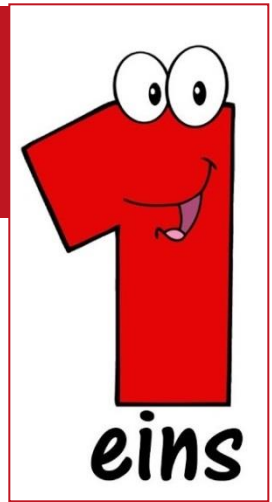
Warum MLM?

Das Basismodell

Erweiterungen des Basismodells

Ein paar allgemeine ‚Musts‘ von MLM

Zusammenfassung



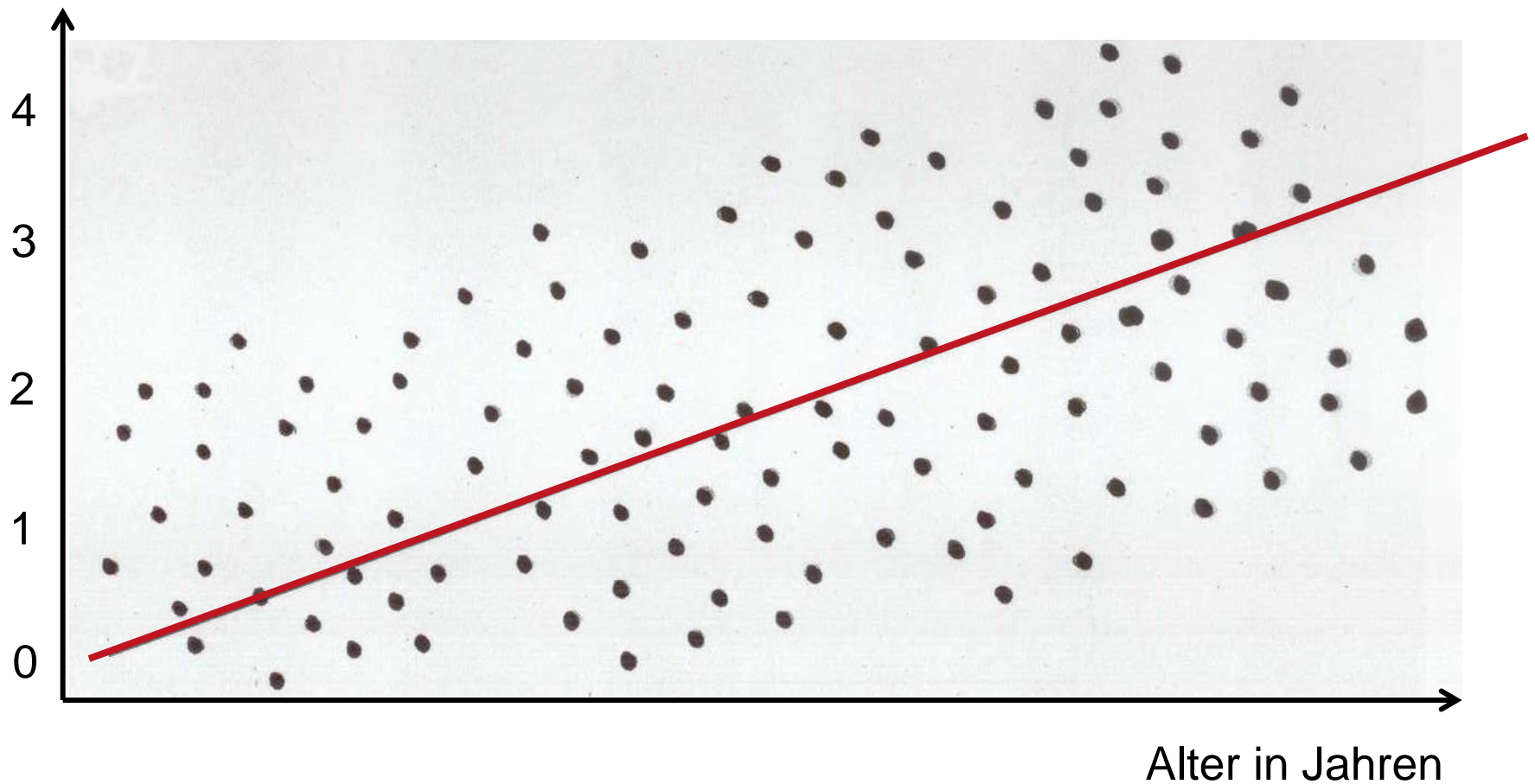
# Warum MLM?



# Ein fiktives Beispiel

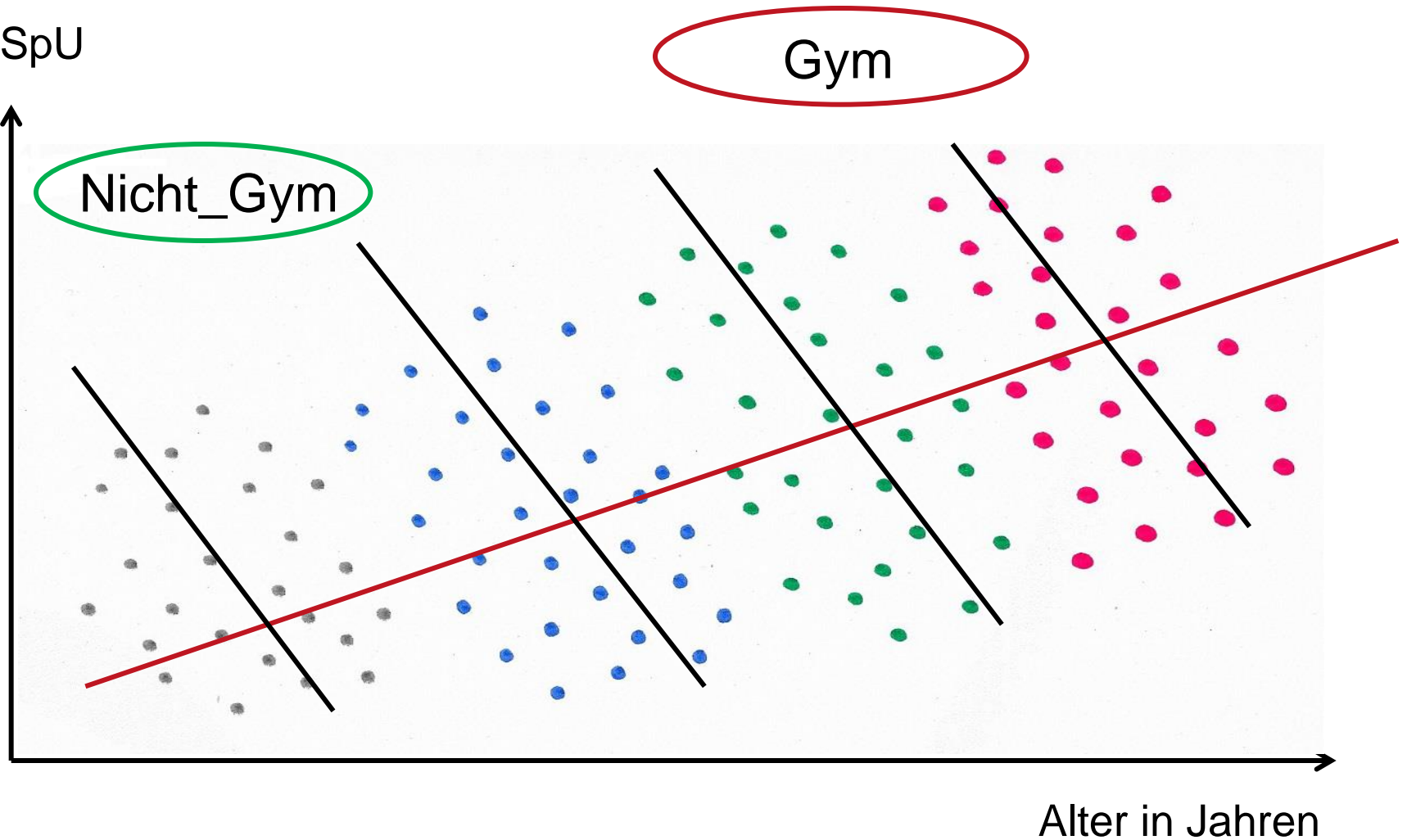
# Das Einebenenmodell

W\_SpU



# Das Zweiebenenmodell

W\_SpU







# Die inhaltliche Perspektive

# Datenstruktur in den Sozialwissenschaften

- Daten aus Studien in den Verhaltens-, Sozial- oder Wirtschaftswissenschaften zeigen sehr häufig hierarchische Strukturen, weil Individuen in größeren Einheiten bzw. Gruppen zusammengefasst sind.
- Beispiele (vgl. Luke, 2004)

	<b>Situation 1</b>	<b>Situation 2</b>	<b>Situation 3</b>
Ebene 1	Schüler	Kinder	Spieler
Ebene 2	Klassen	Familien	Teams
Ebene 3	Schulen	Wohngegend	Sportarten

# Mögliche konzeptuelle Probleme

- Gemessene Werte von Individuen werden letztendlich auch durch die aggregierte Einheit beeinflusst, z. B. das Lernklima (Helmke, 2007) in einer Klasse oder die Sozialstruktur einer Wohngegend (Onwuegbuzie & Hitchcock, 2015).
- Die Missachtung einer hierarchischen Datenstruktur führt zu konzeptuellen Problemen (Hox, 2010, S. 3f), wie etwa
  - einem *ökologischen Fehlschluss* oder
  - einem *atomistischen Fehlschluss* bzw. *Simpson's Paradox*

# Mögliche konzeptuelle Probleme

- Gemessene Werte von Individuen werden letztendlich auch durch die aggregierte Einheit beeinflusst, z. B. das Lernklima (Helmke, 2007) in einer Klasse oder die Sozialstruktur einer Wohngegend (Onwuegbuzie & Hitchcock, 2015).
- Die Missachtung einer hierarchischen Datenstruktur führt zu konzeptuellen Problemen (Hox, 2010, S. 3f), wie etwa
  - einem *ökologischen Fehlschluss* oder
  - einem *atomistischen Fehlschluss* bzw. *Simpson's Paradox*.

# Inhaltliche Zielsetzung eines MLM

- Grundidee MLM ist, den Einfluss einer jeweils höheren Ebene auf die darunter liegende Ebene zu modellieren bzw. zu schätzen.
- In anderen Worten: Ziel von Mehrebenen-Modellierungen ist, die abhängige Variable, die sich auf der untersten Ebene befindet, durch Prädiktoren auf verschiedenen Ebenen zu schätzen.
- Solche Schätzungen sind statistisch durch verschiedene Verfahren möglich, wobei *Regressionsmodelle* und *Kovarianz-Modelle* zentral sind.



# Die statistische Perspektive

# Statistische Argumente

- Traditionelle Methoden, wie z. B. Regressions- oder Varianzanalyse, sind sogenannte Ein-Ebenen-Methoden; das heißt, sie nehmen an, dass Werte von Individuen unabhängig sind.

Verletzung dieser Annahme



Verzerrungen bei statistischen Schätzungen

Zweifelhafte Signifikanzen

- Hauptfehler: **Aggregation & Disaggregation**

# Variablenverschiebung

## Aggregation

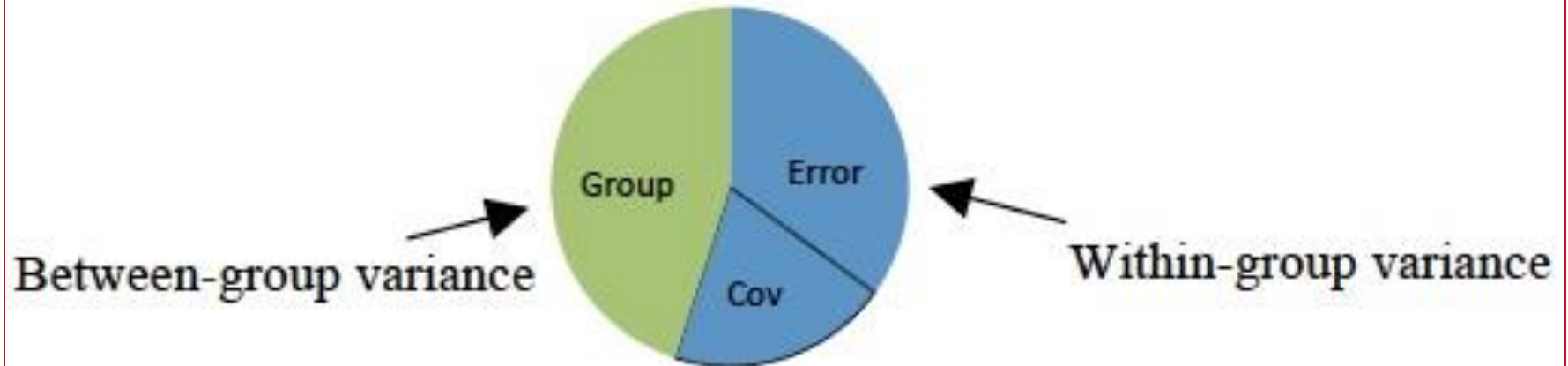
- Die Daten der Ebene 1 werden zusammengefasst und auf der Ebene 2 betrachtet.
- Konsequenzen:
  - ‚Bedeutungsveränderung‘
  - ‚Inhaltliche Fehlschlüsse‘
  - „Cross-level interaction“ wird ausgeschlossen

## Disaggregation

- Die Daten der Ebene 2 werden nicht betrachtet, lediglich Ebene 1 wird ausgewertet.
- Konsequenzen:
  - Stichprobeneffekt wird künstlich erhöht, was zu Verzerrungen führt.
  - ‚Vernichtung‘ der Zusammenhänge



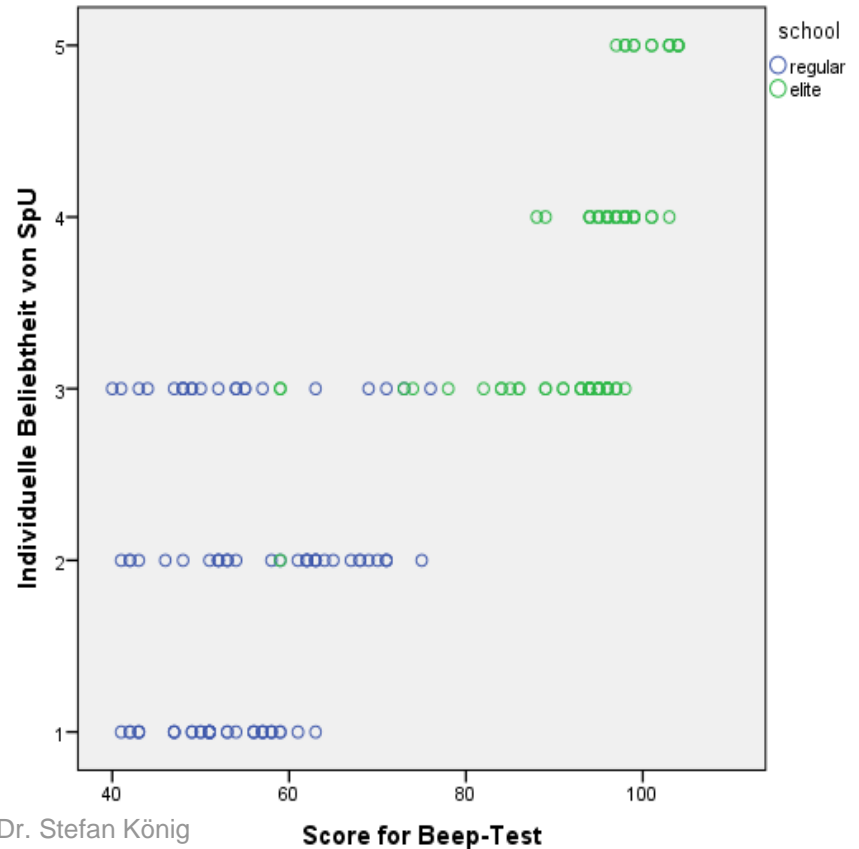
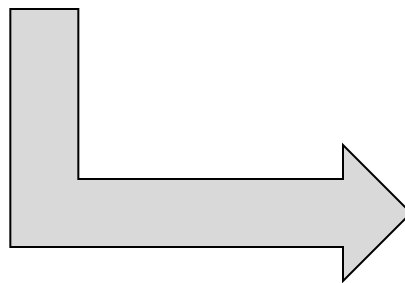
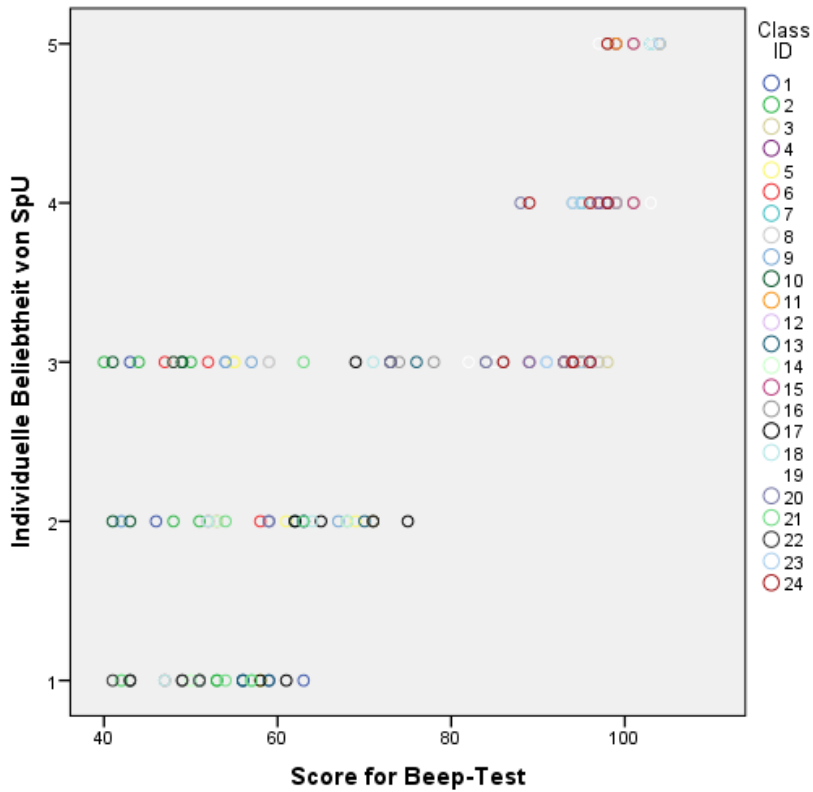
# Grundprinzip der MLMs



Es ist zu prüfen, ob die Varianz zwischen den Gruppen („between“) größer ist als die Varianz innerhalb der Gruppen („within“). Ist dies der Fall, ist ein MLM angezeigt. Prüfung durch ...

- ⇒ ... grafische Datenanalyse oder/und
- ⇒ ... Berechnung des ICC.

# Grafische Datenanalyse



# Der Intraclass Correlation Coefficient (ICC)

- “The ICC (Intraclass Correlation Coefficient) is a measurement of how much correlatedness exists in a hierarchical data set”.
- Definition des Koeffizienten:
  - Varianzanteil der Makroebene (level 2) an der gesamten Varianz der AV.
  - $$\rho = \frac{\textit{between group variance}}{\textit{within group variance} + \textit{between group variance}}$$
- Relevanz des Koeffizienten:
  - A high ICC implies that “the between-group variance dominates the within-group variance. And thus, most of the differences that we see across subjects on Y are actually stemming from group differences. Conversely, a small ICC points to the variance in observed response stems from individual differences within groups (level 2 units)”.
  - Orientierung: 0.05 – 0.12 (Nicht in Stein gemeißelt!!).

# Zwischenfazit

- In den Sozialwissenschaften und Fachdidaktiken liegen häufig geschachtelte Datensätze vor; d. h., die Probanden sind auf einer „höheren“ Ebene aggregiert.
- Das ist nicht bedeutungslos, da etwa Klassen- oder Schulkontext Einstellungen auf Individualebene beeinflussen können.
- MLMs modellieren den Kontexteinfluss, der durch die Gruppe, der ein Individuum angehört, erzeugt wird.



# Das Basismodell

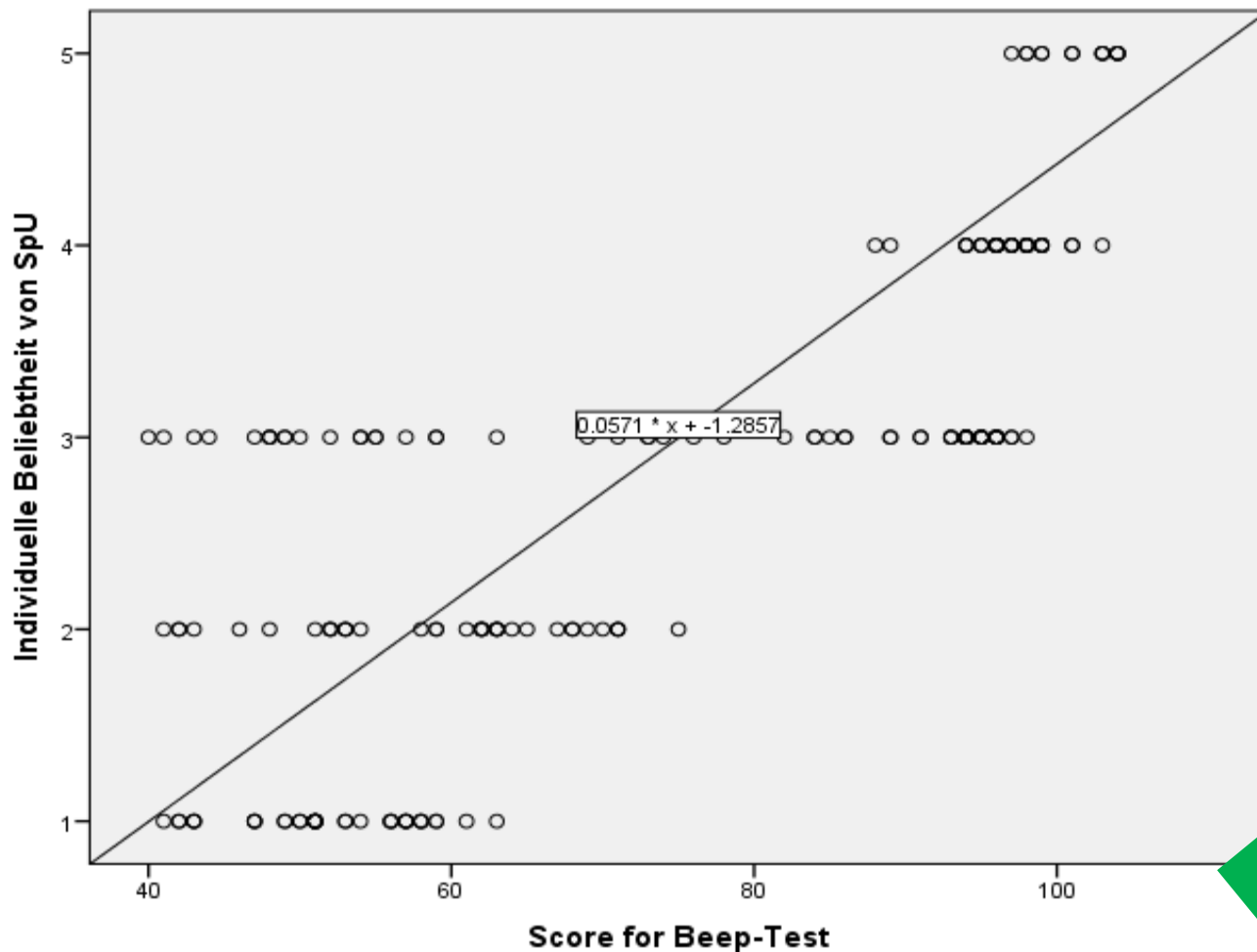
# Beispiel: Ausdauerstudie I

- Forschungskontext: Analyse der Beliebtheit von Sportunterricht in Abhängigkeit individueller Fähigkeiten.
- Forschungsfrage: „*Wie beeinflusst die individuelle körperliche Leistungsfähigkeit (Ausdauer) die individuelle Beliebtheit von Sportunterricht?*“
- Methode:
  - Durchführung der Datenerhebung in mehreren Klassen nach standardisierten Vorgaben.
  - Messung durch sportmotorische Tests und Rating-Skala.
  - Auswertung nach Ein- und Mehrebenen-Verfahren
  - Vergleich

# Datensatz

- Stichprobe:
  - $N = 192$
  - 24 Klassen (0 = normal, 1 = Sportzug)
  - 8 SuS pro Klasse
- Anmerkung zu Stichprobenstruktur
  - Eher kleine Stichprobe für MLM ( $n_g = 192$ ).
  - Wichtig: Eher größere Anzahl an Einheiten auf Ebene 2 ( $n_2 = 24$ ) und nicht unbedingt alle Einheiten auf Ebene 1 ( $n_1 = 8$ ).
- ICC:  $\rho = 0.71$

# Ergebnisse einer OLS-Regression



## Einige Koeffizienten:

- $R^2 = .833$
- $p_{\text{intercept}} = .005$
- $p_{\text{slope}} = .000$

Bitte merken!



# Statistische Mehrebenen-Modellierung

## ■ Ein-Ebenen-Modell (OLS-Regression):

$$Y_i = \beta_0 + \beta_1 x_i + r_i \quad (1)$$

$$Y_i = -1,289 + 0,058x_i + r_i \quad (1)$$

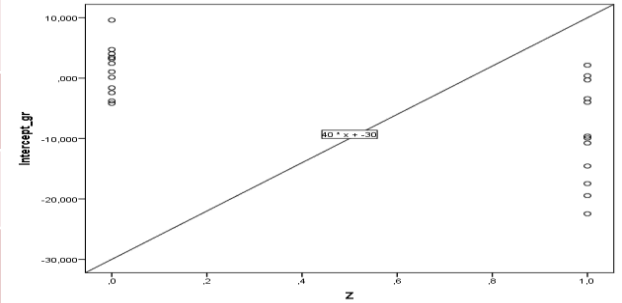
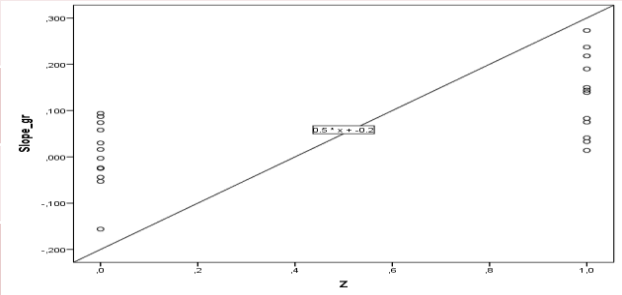
## ■ Zwei-Ebenen-Modell (HLM 2):

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \quad (3)$$

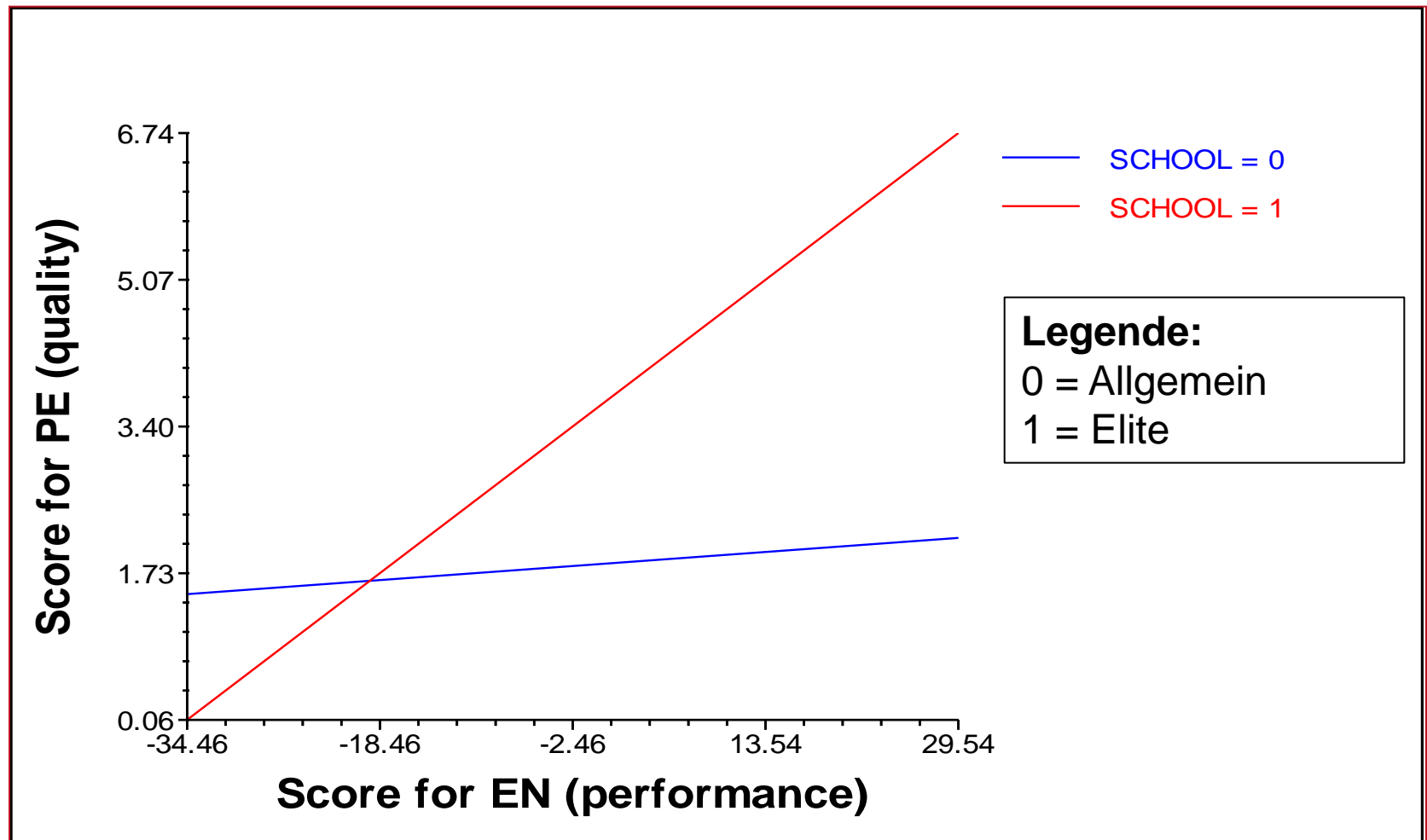
# Erläuterung der Koeffizienten

	Bezeichnung	Bedeutung
$\beta_{0j}$	Ausgangswert ( <i>Intercept</i> ) der Gruppe „j“	Ausgangswert, Durchschnitt
$\beta_{1j}$	Effekt ( <i>Slope</i> ) der Gruppe „j“	Steigung, Effekt
$e_{ij}$	Fehlerterm, Residuum (ind. Abweichung)	Individuelle Abweichung
$\gamma_{00}$	Intercept bzw. slope für Regression aus $\beta_{0j}$ und Z (level 2 dummy)	
$\gamma_{01}$		
$\gamma_{10}$	Intercept bzw. slope für Regression aus $\beta_{1j}$ und Z (level 2 dummy)	
$\gamma_{11}$		
$u_{0j}$	Fehlerterm für Gleichung (2)	
$u_{1j}$	Fehlerterm für Gleichung (3)	

# Ergebnisse eines HLM 2: Koeffizienten

Fixed Effect	Coefficient	Standard error	<i>t</i> -ratio	Approx. <i>d.f.</i>	<i>p</i> -value
For INTRCPT1, $\beta_0$					
INTRCPT2, $\gamma_{00}$	1.833333	0.068041	26.944	22	<0.001
SCHOOL, $\gamma_{01}$	1.822917	0.098225	18.559	22	<0.001
For EN_SCORE slope, $\beta_1$					
INTRCPT2, $\gamma_{10}$	0.010007	0.018507	0.541	22	0.594
SCHOOL, $\gamma_{11}$	0.094485	0.028356	3.332	22	0.003

# Ergebnisse eines HLM 2: Graph



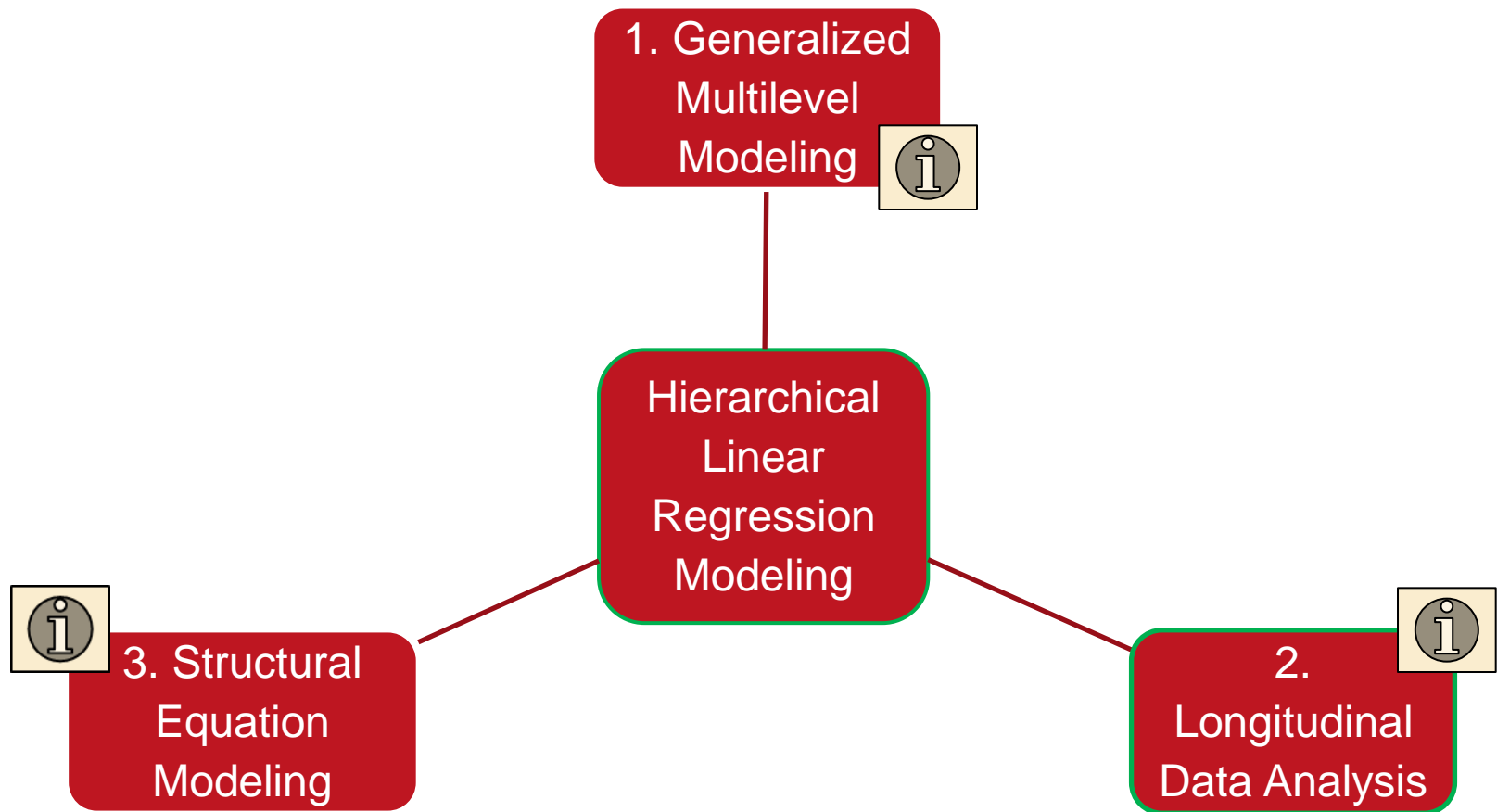
# Zwischenfazit

- Daten in den Sozial- und Verhaltenswissenschaften sind sehr häufig „geschachtelt“.
- Eine Anwendung von „Ein-Ebenen-Methoden“, deren Einsatz auf Aggregation bzw. Disaggregation der Daten beruht, führt zu Verzerrungen.
- Als Konsequenz sind bei der statistischen Auswertung deshalb MLM zu fordern, falls der ICC einen bestimmten Level überschreitet.
- Konkret wurde dies an einem Basismodell gezeigt; im Folgenden möchte ich dieses Basismodell modifizieren bzw. erweitern.



# Erweiterungen des Basismodells

# Multilevel Modeling – ein Überblick





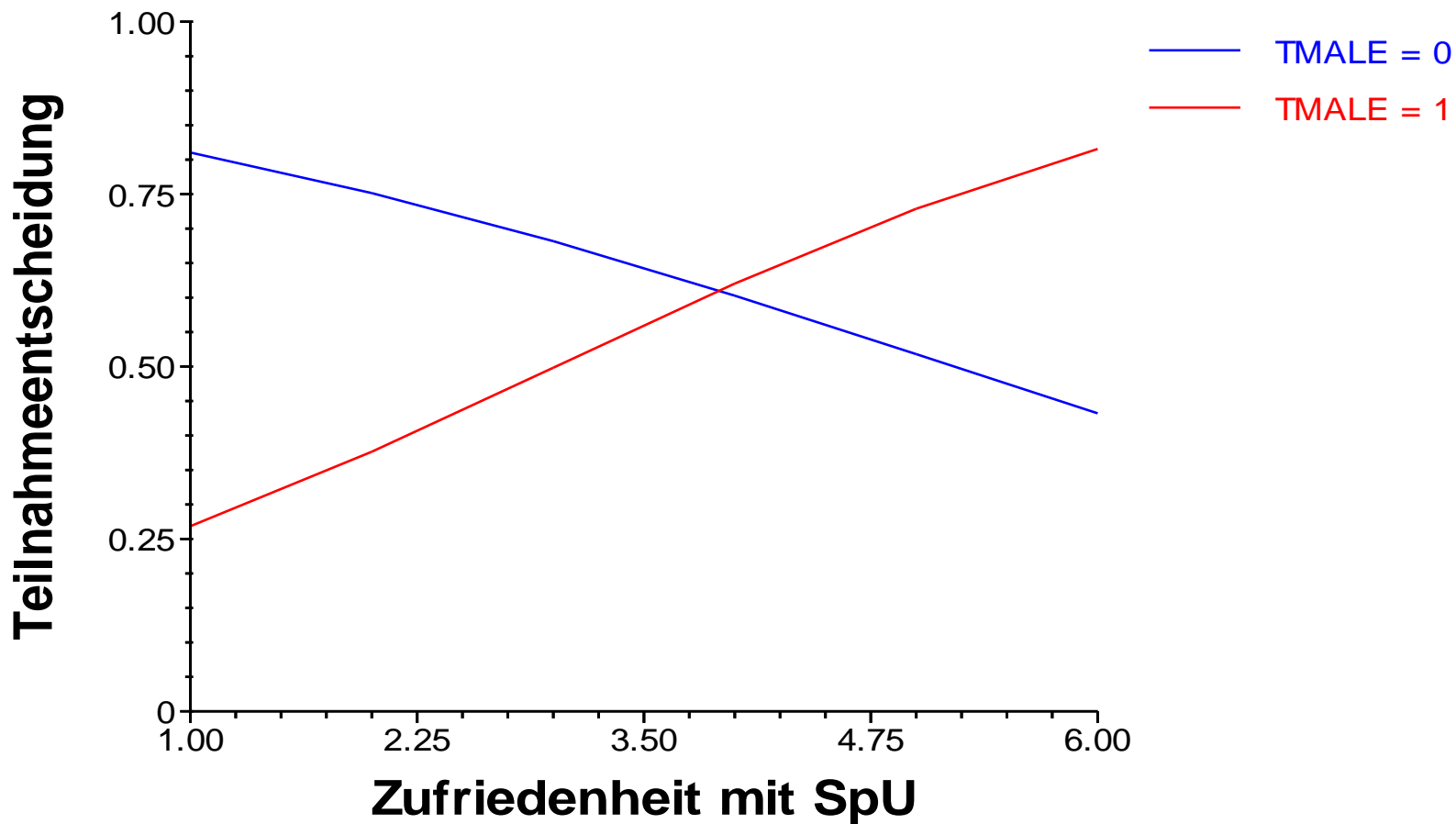
# 1. Erweiterung: GHLM




# GHLM: Prognose dichotomer Variablen

- In den **Sozial- und Verhaltenswissenschaften** müssen in vielfältigen Situationen und Kontexten Entscheidungen „Ja/Nein“ modelliert werden.
- Beispiele:
  - Einschulung Migrantenkinder (Tuppat & Becker, 2014)
  - Berufliche Umzugsentscheidungen (Auspurg et al., 2014)
- Diese Grundidee kann auch auf **sportwissenschaftliche Fragestellungen** übertragen werden.
  - Beispiele Kadermitgliedschaft Nationalmannschaft (Höner et al., 2012)
  - Teilnahme an außerunterrichtlichen Sportprogrammen
- Bei solchen Fragen stößt eine lineare Modellierung (Basismodell) eindeutig an ihre Grenzen!

# Beispiel: Teilnahme an außerunterrichtlichem Sportprogramm





## 2. Erweiterung: Längsschnittdaten



**Methodologie**



**Beispiel**



**Zusammenfassung**

# (1) Methodologie des GHLM

- In den **Sozial- und Verhaltenswissenschaften** wird oft versucht, individuelle Entwicklungsverläufe („Change over time“) in Abhängigkeit von theoretisch sinnvollen Prädiktoren abzubilden.
- Beispiele:
  - Wirkungen von Verhaltenstherapien (Göllner et al., 2010)
  - Analyse von „Aging“ (cf. Luke, 2004)
- Diese Grundidee kann auch auf **trainingswissenschaftliche Fragestellungen** übertragen werden.
  - Schulleistung bei Kaderathleten (Wartenberg et al., 2014)
  - Wirkungen von verschiedenen Trainingsmethoden
- Einzelne Beobachtungen gelten als „nested within a subject“

# Statistische Betrachtung

## rANOVA

- Der rANOVA-Ansatz kann solche Fragestellungen modellieren, hat aber folgende Nachteile:
  - **Fälle mit fehlenden Werten** werden von SPSS ausgeschlossen => geringere Power
  - **Messzeitpunkte** werden als zeitgleich behandelt, was nicht grundsätzlich stimmt.

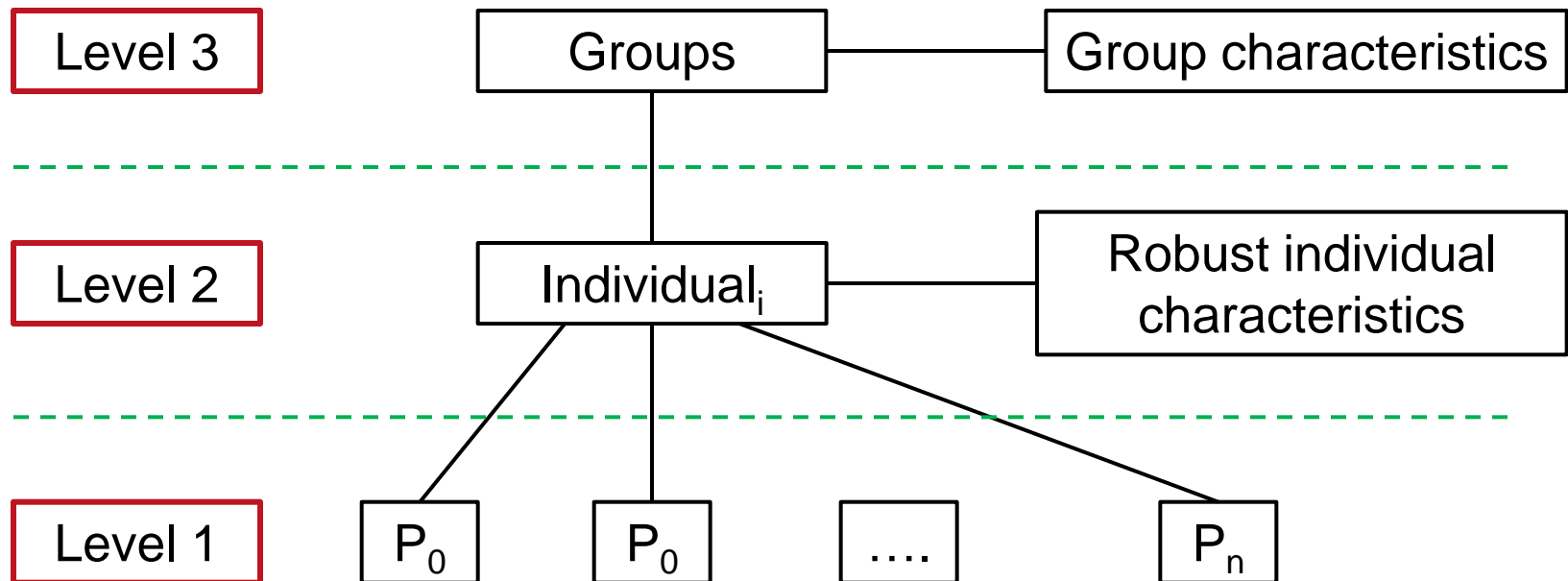
## HLM

- HLM basiert auf OLS-Regressionen, also linearen Funktionen. Das hat folgende Vorteile:
  - Fehlende Werte spielen kaum eine Rolle
  - Dto. Messzeitpunkte
  - Kontextbedingungen werden besser modelliert, was ein quasi-experimentelles Design besser abbildet

# Drei wichtige Merkmale einer Längsschnittstudie (vgl. Singer & Willett, 2003, pp. 9–15)

- Drei oder mehr Datenerhebungen (**Messzeitpunkte**)
  - Bitte echte Längsschnitterhebungen durchführen und nicht unterschiedliche Teilkohorten einer Querschnittsstudie vergleichen!
  - Zwei Messzeitpunkte können eine individuelle Entwicklung nicht gut abbilden und berücksichtigen keine Messfehler.
- Ein **Messergebnis (AV)**, das sich systematisch über die Zeit verändert.
- Vernünftiger **Messzeit-Takt**, der eine inhaltliche Begründung erfordert (z. B. Trainingsprinzip vs. Medikamentenwirkung).

# Drei-Ebenen-Modell: Modellierung



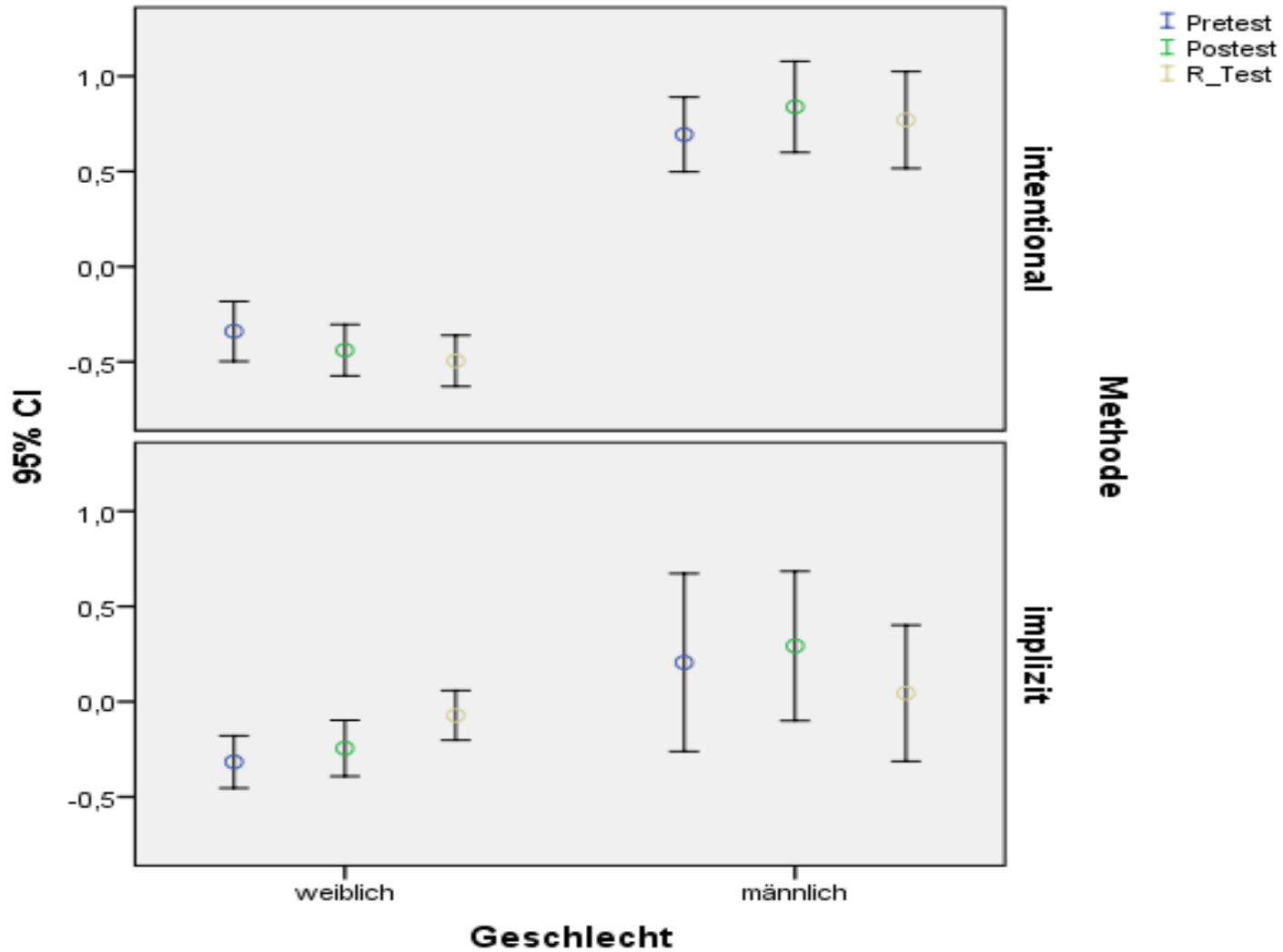
- Level-3: Groups (Snijders & Bosker, 2010, p. 247)
- Level-2: Robust individual characteristics (Singer & Willett, 2003, pp. 57–63)
- Level-1: Measurement points (Luke, 2004, pp. 62–64)



## (2) Beispiel

- Forschungskontext: Analyse trainingswissenschaftlicher Prozesse im Rahmen sportdidaktischer Forschung
- Forschungsfrage: „Führen intentionale und implizite Krafttrainingsprogramme im Sportunterricht zu messbaren Effekten bei Schüler\*innen?“
- Methode:
  - Datenerhebung mit sportmotorischen Tests (6 Krafttests) zu 3 MZP
  - Aufbereitung der Daten und Berechnen eines Gesamtscores mittels z-Transformation
  - Deskriptive und inferenzstatistische Auswertung

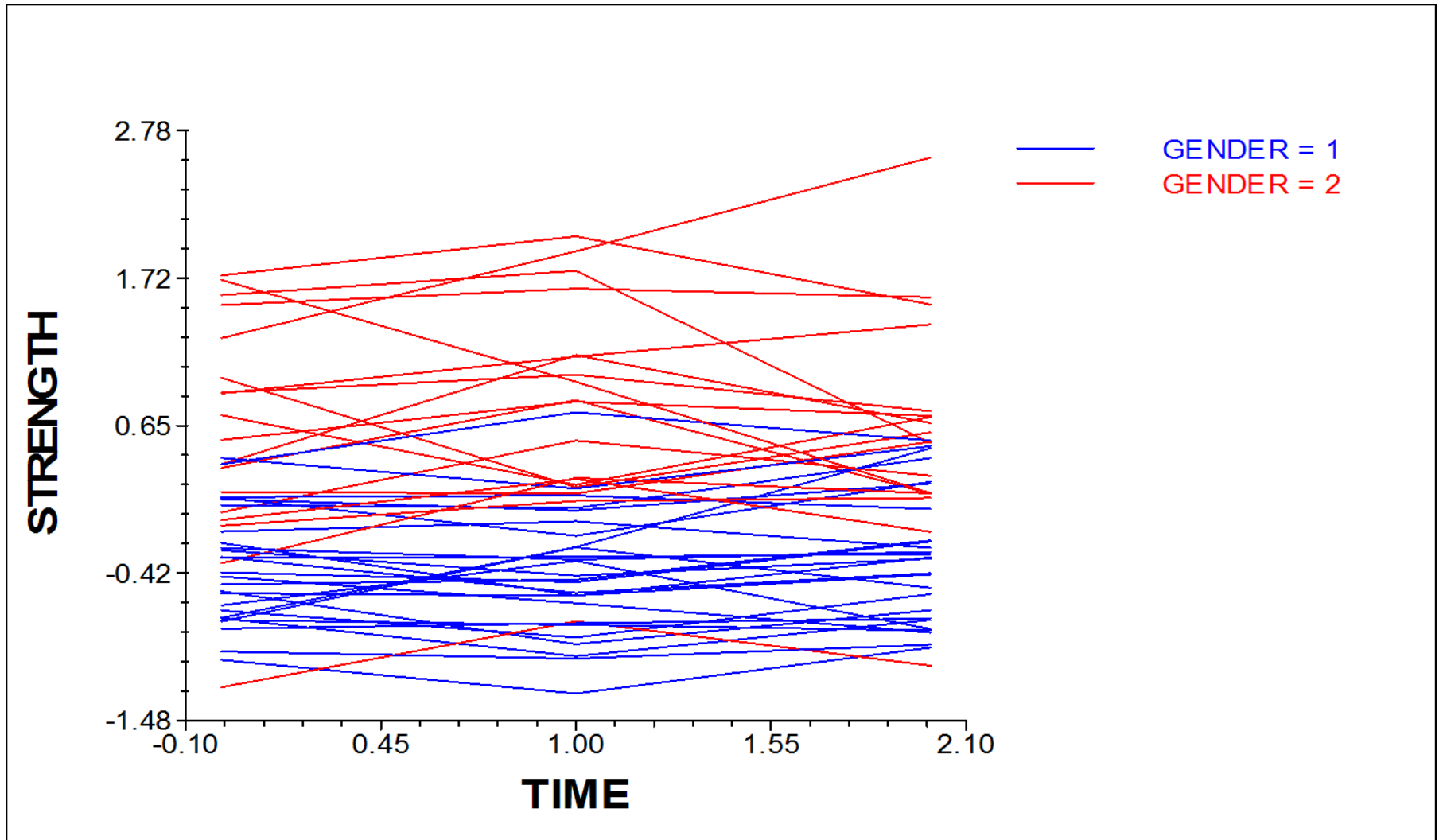
# ANOVA-Ansatz



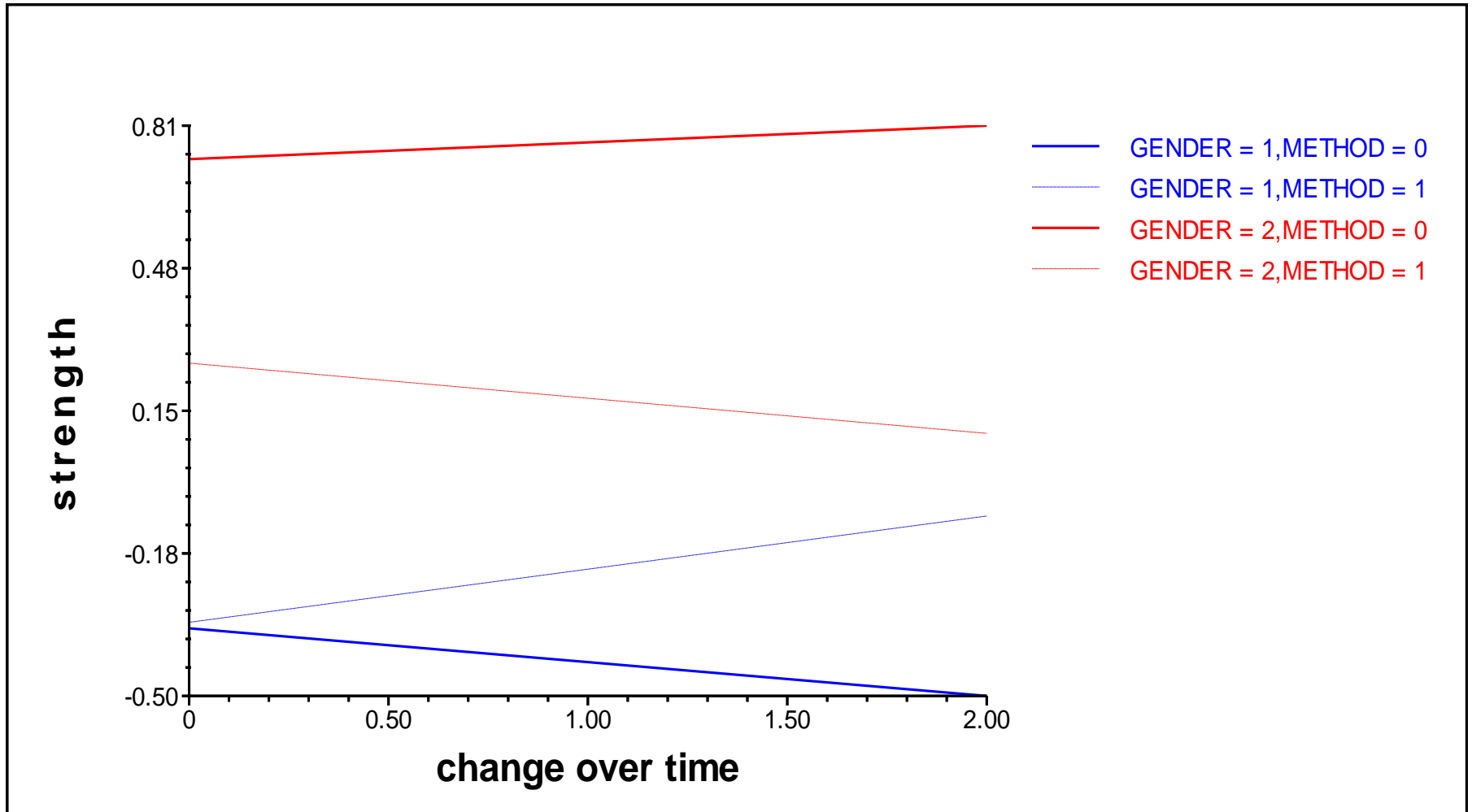
# HLM 3: Koeffizienten

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
<b>For INTRCEPT1, <math>\pi_0</math></b> (starting value of individual i at time = 0 [baseline])					
For INTRCEPT2, $\beta_{00}$ (average initial status for level 2 = 1 [female])					
INTRCEPT3, $\gamma_{000}$	-1.423216	0.203483	-6.994	137	0.002
METHOD, $\gamma_{001}$	0.495575	0.290333	1.707	137	0.163
For GENDER, $\beta_{01}$ (difference in initial status between female and male participants)					
INTRCEPT3, $\gamma_{010}$	1.076207	0.134735	7.988	137	<0.001
METHOD, $\gamma_{011}$	-0.481766	0.208632	-2.309	137	0.022
<b>For TIME slope, <math>\pi_1</math></b> (rate of change of individual i [effect])					
For INTRCEPT2, $\beta_{10}$ (average change for level 2 = 1 [female])					
INTRCEPT3, $\gamma_{100}$	-0.193473	0.074259	-2.605	137	0.010
METHOD, $\gamma_{101}$	0.517990	0.106011	4.886	137	<0.001
For GENDER, $\beta_{11}$ (difference in change between female and male)					
INTRCEPT3, $\gamma_{110}$	0.115913	0.049179	2.357	137	0.020
METHOD, $\gamma_{111}$	-0.318435	0.076205	-4.179	137	<0.001

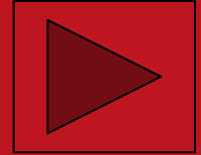
# HLM 3: Individuelle Entwicklungsverläufe



# HLM 3: Graph von Level 2 und 3



# (3) Zwischenfazit



- Daten in den Sozial- und Verhaltenswissenschaften sind häufig auch als *Längsschnitte* oder *Wachstumskurven* angelegt.
- Überträgt man die Idee von MLM auf diese Situation, dann werden einzelne Messungen auf Ebene 1 einem Individuum auf Ebene 2 zugeordnet.
- Dieser Ansatz hat gegenüber dem eher traditionellen ANCOVAR-Ansatz gewisse Vorteile, wird aber durch eine häufig notwendige dritte Ebene relativ komplex.
- Dennoch ermöglicht dieser Ansatz ebenfalls sehr viel robustere Schätzungen, zumal durch die Integration sogenannter „random effects“ individuelle Entwicklungen sehr viel mehr in den Blickpunkt genommen werden.
- Als mächtige Alternative stehen den MLM seit einiger Zeit die LGCMs gegenüber.



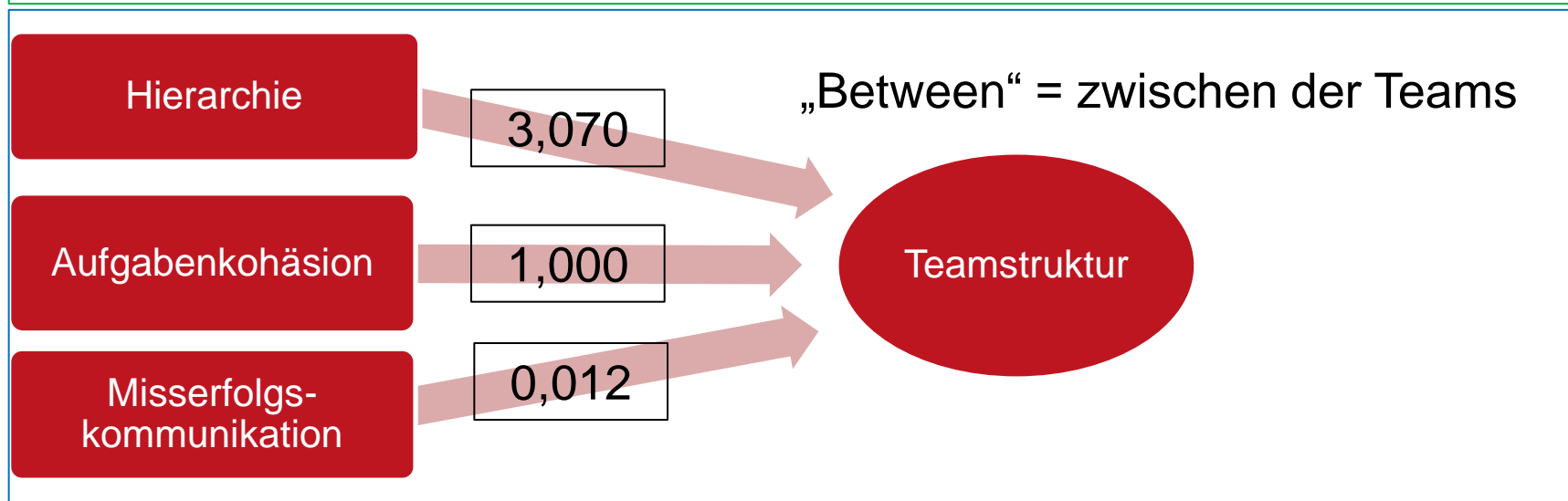
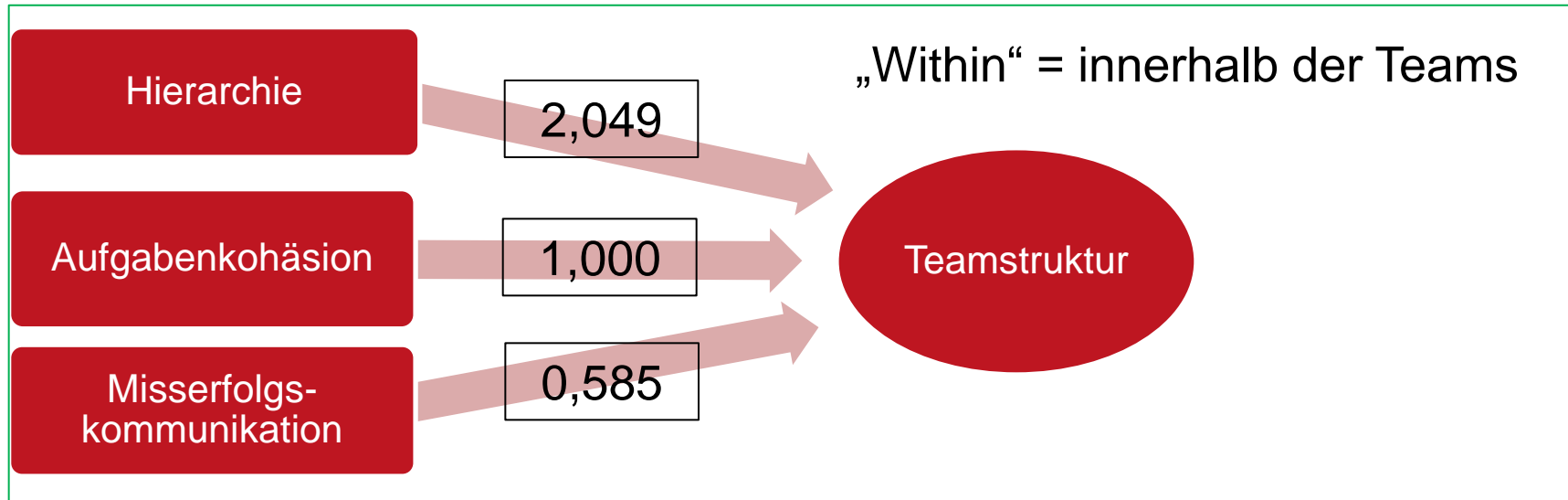
# 3. Erweiterung: KMFA

# KMFA: Latente Konstrukte

- In den **Sozial- und Verhaltenswissenschaften** wird oft versucht, individuelle Entwicklungsverläufe („Change over time“) in Abhängigkeit von theoretisch sinnvollen Prädiktoren abzubilden.
- Beispiele:
  - Anstrengung bei Hausaufgaben (Lüdtke et al., 2007)
  - Arbeitsplatzzufriedenheit (cf. Luke, 2004)
- Diese Grundidee kann auch auf **trainingswissenschaftliche Fragestellungen** übertragen werden.
  - Qualität von Sportunterricht (Wartenberg et al., 2014)
  - Belastungsempfinden von Sportlehrkräften (Oravec i. Vb.)
- Einzelne Beobachtungen gelten als „nested within a subject“.



# Beispiel: Mannschaftsführung





# Ein paar generelle ‚Musts‘ bei MLM

```
graph TD; A[Theoretische Modellentscheidungen] --> B[Stichprobenziehung]; B --> C[Variablen]; C --> D[Das ‚best‘ Modell];
```

Theoretische  
Modellentscheidungen

Stichprobenziehung

Variablen

Das ‚best‘ Modell

# (1) Modellentscheidungen

- Wie viele Ebenen sind in den Daten repräsentiert?
  - Die meisten Studien beziehen sich auf 2 oder 3 Ebenen, 4 sind derzeit modellierbar.
- Wie viele Prädiktoren sollen pro Ebene betrachtet werden?
- Sollen *intercepts* oder *slopes* oder beides auf Ebene 1 modelliert werden?
- Welche Teile des Modells sollen *random effects* beinhalten?

## (2) Stichprobenziehung

- Grundsätzlich erfordern Mehrebenenmodelle größere bis große Stichproben, um eine entsprechende Power zu realisieren.
- Allerdings erfordert MML – wie andere komplexe Verfahren auch (Kline, 2005) – auch spezifische Überlegungen zur Stichprobenziehung.
- Empfehlung (u. a. Maas & Hox, 2005; Hox, 2010; Snijders & Bosker, 2012; McNeish & Stapleton, 2016): „Da die Stichprobengröße auf der höchsten Ebene die limitierende Größe darstellt, sollte sie eher groß sein ( $\geq 30$  Kontexte); als Konsequenz, muss auf den Ebenen darunter nicht jedes Gruppenmitglied getestet werden“.

## (3) Variablen

- Negativbeispiel „Qualitätsmerkmale von Sportunterricht“
  - *Ebene 1*: Variablen zu motivationalen und volitionalen Faktoren sowie zur Unterrichtsqualität
  - *Ebene 2*: Aggregation dieser Daten zu einem Klassenmittelwert
- Entscheidend ist, dass auf den entsprechenden Ebenen mit den spezifischen Variablen gearbeitet wird, d. h.
  - Individualdaten auf Ebene 1
  - Gruppenmerkmalen auf Ebene 2
- Schließlich ist bei der Bezeichnung der Variablen darauf zu achten, dass die Ebenenzugehörigkeit erkennbar ist.

## (4) Das ‚best‘ Modell

- Bei Mehrebenenanalysen ist es nicht damit getan, ein Modell zu berechnen und dann zu glauben, ...“wow!“
- Vielmehr ist es Standard, mehrere konkurrierende Modelle mittels sogenannter Fit-Indizes zu prüfen und zu vergleichen. Die bekanntesten sind die Loglikelihood-Funktion ( $-2LL$ ), das AIC und das BIC; sie werden als *absolute Fit Indizes* bezeichnet.
- Probleme:
  - Nicht alle Software-Pakete berechnen diese Koeffizienten automatisch
  - *Modell Over-fitting* beachten!.

# Zusammenfassung



- Daten in den Sozial- und Verhaltenswissenschaften sind häufig geschachtelt, weswegen von Kontexteffekten auszugehen ist.
- MLMs modellieren diesen Kontexteinfluss, der durch die Gruppe, der ein Individuum angehört, erzeugt wird.
- Bei der Modellierung ist zu beachten, dass auf Level 1 nur Individualvariablen und auf Level 2 nur Gruppenvariablen verwendet werden.
- Grundsätzlich sollte niemals nur ein Modell gebildet werden, sondern immer mehrere, um das sogenannte „best“-Modell auswählen zu können.

# *Danke für die Aufmerksamkeit*



# Literatur

- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied Longitudinal Analysis*, 2nd ed. New Jersey: WILEY.
- Hox, J. J. (2010). *Multilevel Analysis – Techniques and Applications*. New York and Hove: Routledge.
- Luke, D. A. (2004). *Multilevel modeling* (Sage University Paper Series on Quantitative Applications in Social Sciences, 07-143). Newbury Park, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. T., Congdon, R., & Toit, M. (2011). *HLM 7. Hierarchical Linear and Nonlinear Modelling*. Lincolnwood: Scientific Software International.
- Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. New York: Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.