

Lineare Gemischte Modelle (1)



We are happy to share our materials openly:

The content of these [Open Educational Resources](#) by [Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München](#) is licensed under [CC BY-SA 4.0](#). The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

Zur Erinnerung:

- Das Standardmodell der multiplen Regression nach dem Least-Squares-Ansatz lautet:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i$$

- Das heißt, jeder beobachtete Wert y einer Person i setzt sich zusammen aus einer Konstanten β_0 plus der Summe aus k mit dem Faktor β_k gewichteten Einzelwerten X_{ki} und einem Zufallsfehler ϵ_i .
- Für die Durchführung statistischer Tests wird weiterhin angenommen, dass dieser Zufallsfehler normalverteilt mit Erwartungswert 0 und konstanter (d.h. homoskedastischer) Varianz ist:

$$\epsilon_i \sim N(0, \sigma_\epsilon^2)$$

Weitere, für alles Folgende sehr wesentliche, Annahme:

➤ **Unabhängigkeit der Fehler ε_i**

identically and independently
distributed

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

- D.h. die Residuen r_i dürfen nicht **systematisch** mit einem Merkmal zusammenhängen.
- Unabhängigkeit konkret ausgedrückt: Falls Unabhängigkeit gegeben ist, dann beinhaltet der Wert einer Person keine Informationen über den Wert einer anderen Person
→ *Gegenbeispiele*: Abschauen bei Statistik-Klausur; Schüler*in einer Klasse mit derselben Lehrkraft (unterliegen gleichem Einfluss → Testergebnisse daher ähnlicher als zu Schüler*innen anderer Klassen).

Unabhängigkeit der Beobachtungen bzw. Fehler ist aber selten gegeben:

- **Abhängigkeit kann immer dann vorliegen, wenn mehrere Beobachtungen zu einer sinnvoll definierten „Gruppe“ oder „Cluster“ gehören**

- **Beispiele:**
 - Schüler*innen sind Schulen zugeordnet
 - Patient*innen sind Therapeut*innen zugeordnet
 - Messzeitpunkte sind Personen zugeordnet

Häufig liegt mehr als eine Gruppierungsstruktur vor. Solche Fällen können nochmal unterteilt werden in:

➤ **Hierarchische Datenstrukturen („hierarchical“):**

- Patient*innen bei Therapeut*innen in Therapiezentren:
 - Jede Patient*in ist genau einer Therapeut*in zugeordnet
 - Jede Therapeut*in ist genau einem Therapiezentrum zugeordnet

➤ **Überlappende Datenstrukturen („cross-classified“):**

- Schüler*innen in Klassen und Stadtvierteln:
 - Nicht alle Schüler*innen aus einem Stadtviertel gehen in die selbe Klasse
 - Nicht alle Schüler*innen aus einer Klasse wohnen im selben Stadtviertel
- In dieser Vorlesung werden wir uns hauptsächlich mit hierarchischen Beispielen beschäftigen. Der Aufbau der statistischen Modelle ist aber meist in beiden Fällen identisch. → „**Multilevel Modelle**“

Gruppierte Datenstruktur:

- Daten mit mehreren Ebenen, den sog. Levels.
- Auf jeder Ebene gibt es klar definierte Elemente.
- Level 1 („L1“): z.B. Individualebene (Schulleistungstestwerte).
- Level 2 („L2“): z.B. Gruppenebene (Schule, Klinik, Abteilung).
- Optional: Weitere Gruppierungsstruktur (Firma, Klinikverbünde, Bundesländer, Nationen), entweder überlappend (weitere „L2“ Variable) oder hierarchisch (Level 3, „L3“ Variable)

Datenlage:

- Die **abhängige Variable (Y)** wird immer auf der **untersten Ebene**, Level 1, gemessen.
 - Bsp. Schulleistungstestergebnis
- **unabhängige Variablen (X_1, X_2, X_{\dots}, X_k)** können auf **allen Ebenen** erhoben werden
 - Bsp. Level 1: IQ-Werte; Level 2: Klassengröße
- Gruppen auf den unterschiedlichen Ebenen können unterschiedlich groß sein

- Personen auf L1, geschachtelt in Gruppen:

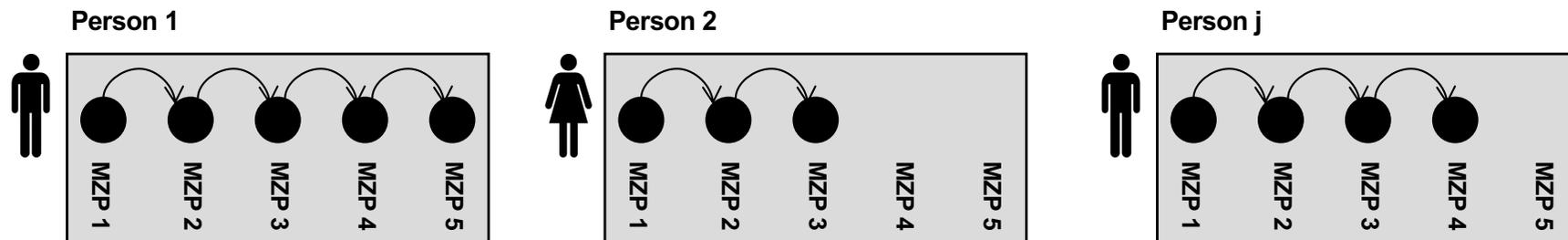
**L2-Prädiktoren: Größe der Klasse, Durchschnittseinkommen,
privat/staatliche Schule, Lehrerpersönlichkeit**



L1-Prädiktoren: Alter, IQ, Geschlecht

- Personen als L2-Struktur / „persons-as-context“:

L2-Prädiktoren (Personenebene): IQ, Alter, Geschlecht



L1-Prädiktoren (Messzeitpunktebene): Tageszeit, (state)-Motivation, ...

Implikationen gruppierter Datenstrukturen:

- Personen innerhalb eines Clusters sind sich **ähnlicher**, weil sie einem gemeinsamen Einfluss ausgesetzt sind.
- Personenmesswerte sind daher voneinander **abhängig**.

Folgen der Vernachlässigung von Abhängigkeiten:

- **Verletzung einer zentralen Annahme der einfachen/multiplen Regression.**
- **Falscher Standardfehler/Prüfverteilung der Regressionskoeffizienten für Signifikanztests wird verwendet**
- **Alpha-Inflation, da die Nullhypothese zu häufig fälschlich abgelehnt wird.**

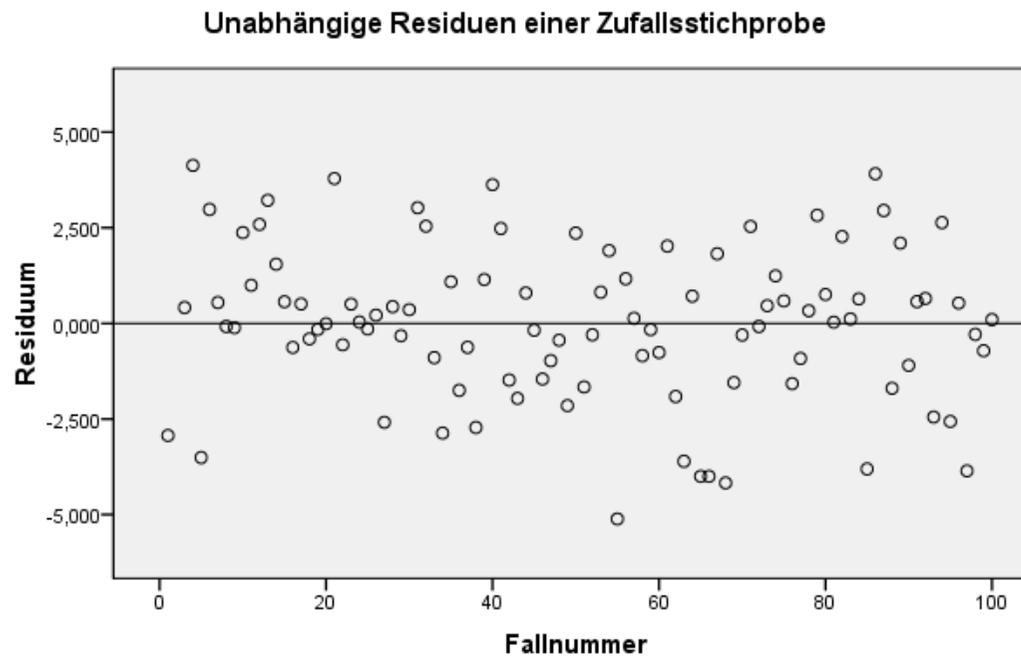
Wie lässt sich Abhängigkeit in den Fehlern darstellen?

Zwei einfache Möglichkeiten:

1. Grafisch-deskriptiv

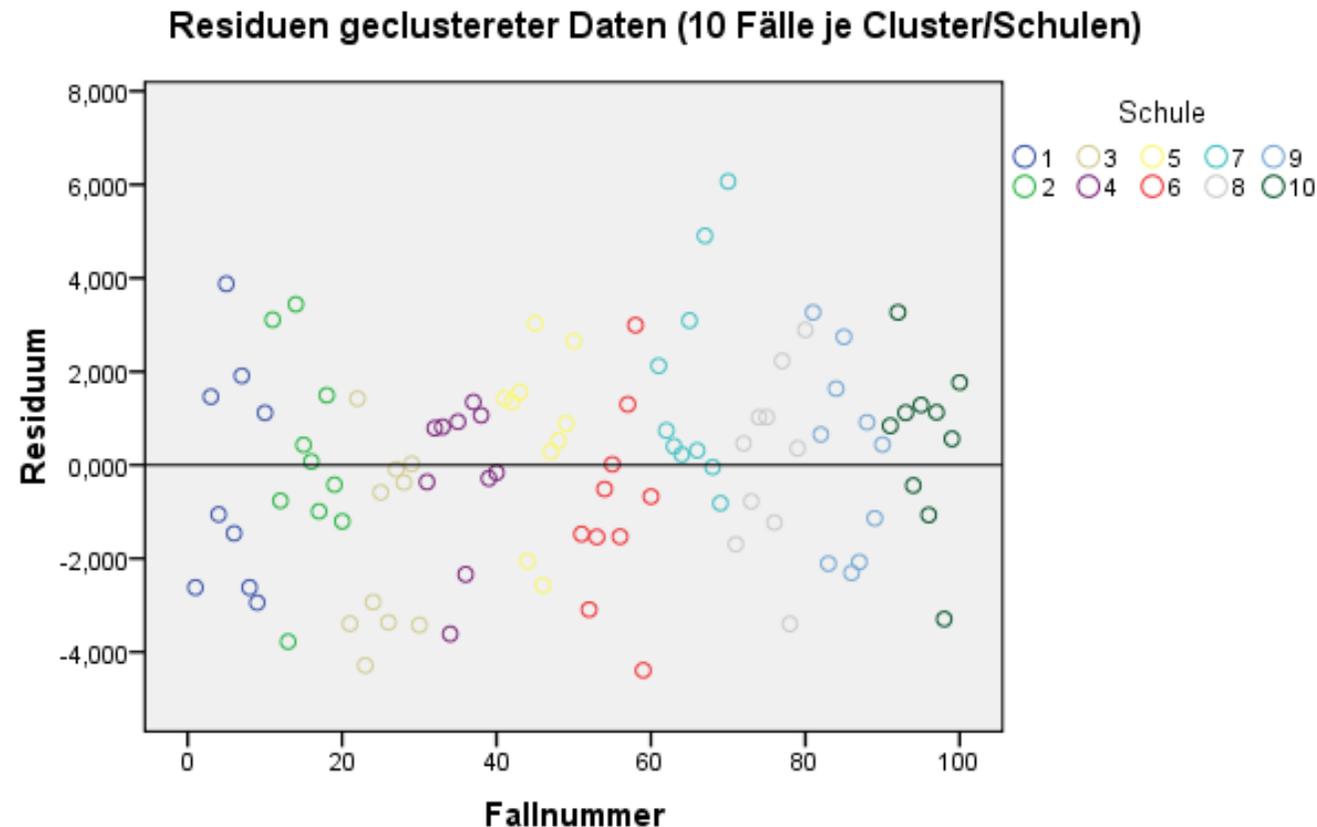
2. Intraklassenkorrelation (ICC)

Residuen-Plot



Über alle Fälle hinweg
betrachtet streuen die
Residuen *unsystematisch* um
den Erwartungswert →
Unabhängigkeit gegeben

Residuen-Plot



Zufällig gezogene
Beobachtungen
stammen aus 10
Schulen: Residuen
innerhalb einer Schule
bzw. eines Clusters
tendieren stärker dazu,
enger beieinander zu
liegen als die Residuen
einer reinen
Zufallsstichprobe

→ Unabhängigkeit der
 ε_i nicht gegeben

Wie lässt sich Abhängigkeit in den Fehlern darstellen?

Zwei einfache Möglichkeiten:

1. Grafisch-deskriptiv

2. Intraklassenkorrelation (ICC)

→ Wie lässt sich Ausmaß der Abhängigkeit in einem Wert erfassen?

$$ICC = \frac{\sigma_{\text{Level-2}}^2}{\sigma_{\text{Gesamt}}^2} = \frac{\sigma_{\text{Level-2}}^2}{\sigma_{\text{Level-2}}^2 + \sigma_{\text{Level-1}}^2}$$

Mit:

$\sigma_{\text{Level-2}}^2$ = Schätzer der Varianz *zwischen* den Gruppen: **Effekt der Gruppenzugehörigkeit.**

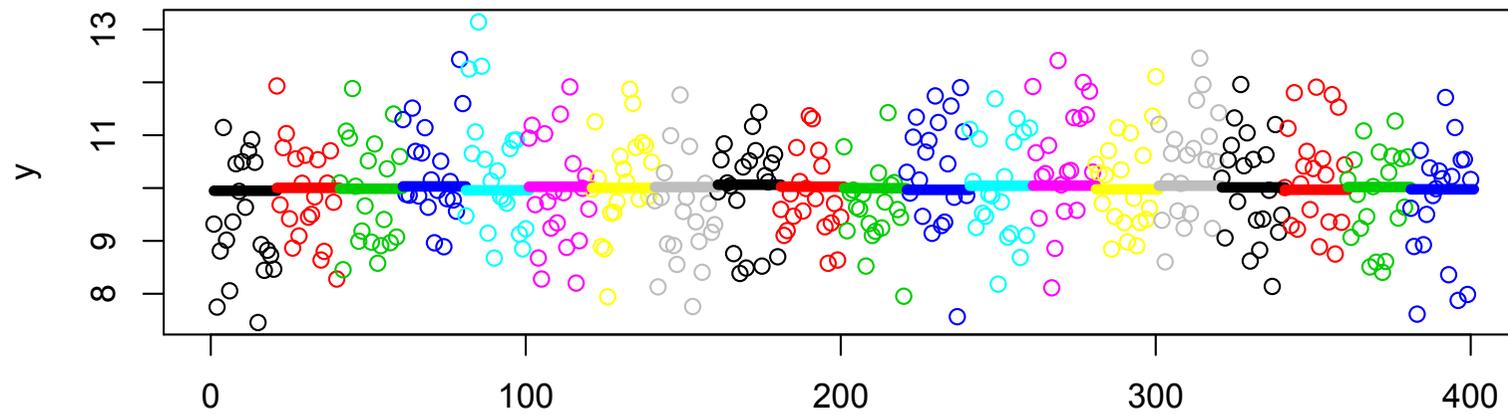
$\sigma_{\text{Level-1}}^2$ = Schätzer der Varianz *innerhalb* der Gruppen.

$\sigma_{\text{Level-2}}^2 + \sigma_{\text{Level-1}}^2$ = Schätzer der *Gesamtvarianz*.

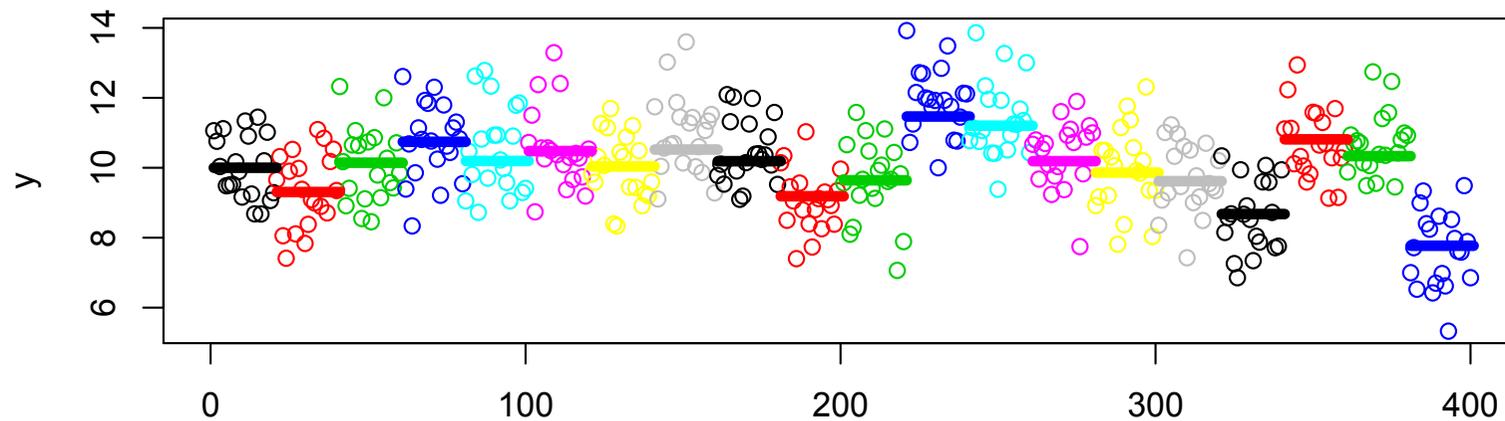
- $\sigma_{\text{Level-2}}^2 = 0 \Rightarrow ICC = 0 \Rightarrow$ **es liegt kein** Gruppierungseffekt vor. Die Gesamtvarianz der abhängigen Variablen geht vollständig auf die Level-1 Varianz zurück. Zwei Personen *einer* Gruppe sind sich nicht ähnlicher als zwei Personen aus *zwei verschiedenen* Gruppen.
- $\sigma_{\text{Level-2}}^2 > 0 \Rightarrow ICC > 0 \Rightarrow$ **es liegt ein** Gruppierungseffekt vor. Zwei Personen *einer* Gruppe sind sich ähnlicher als zwei Personen aus *zwei verschiedenen* Gruppen.
 - ICC gibt also den Prozentsatz an, zu dem die Gesamtvariation der abhängigen Variablen auf die Gruppenzugehörigkeit zurückzuführen ist.

20 Gruppen mit je 20 Personen

ICC = 0%

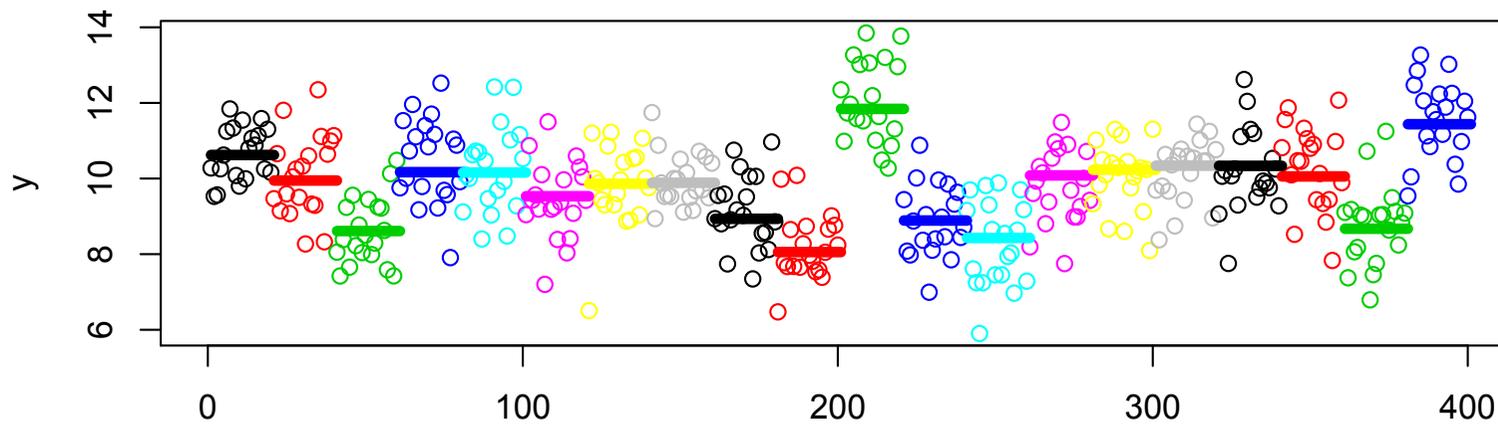


ICC = 29%

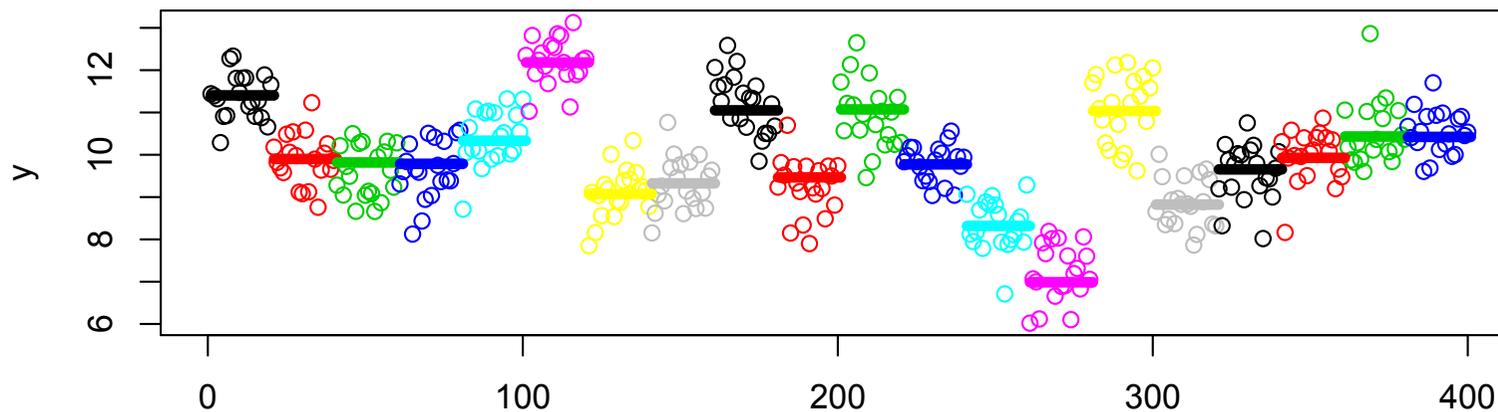


20 Gruppen mit je 20 Personen

ICC = 50%

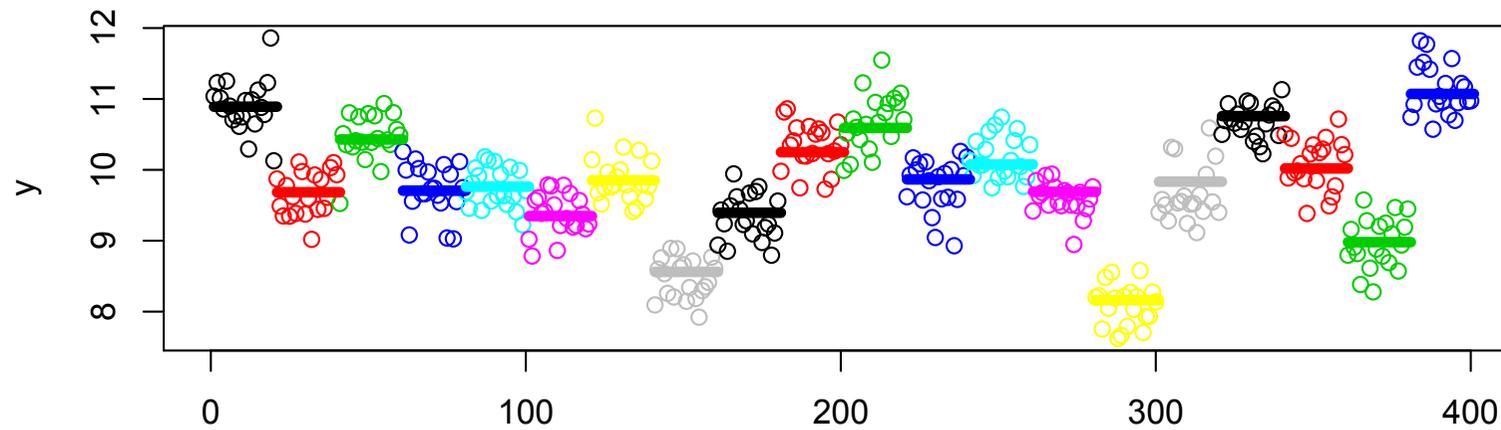


ICC = 71%

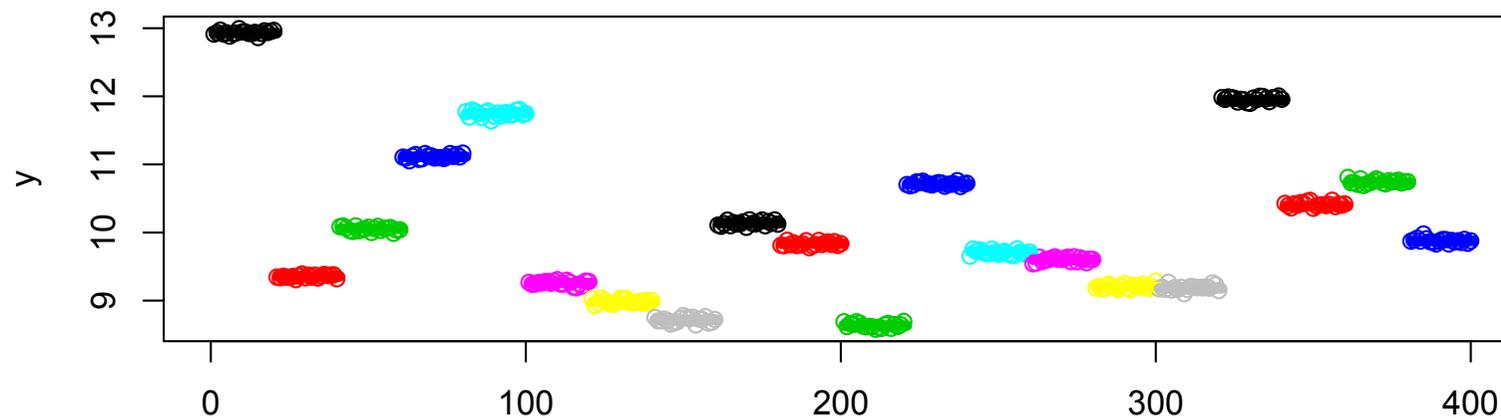


20 Gruppen mit je 20 Personen

ICC = 91%



ICC = 100%



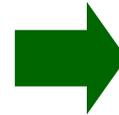
- Wie könnte man nun mit herkömmlichen regressionsanalytischen Methoden gruppierte Daten analysieren?
 - **Aggregation der Daten**
(d.h. nur auf Gruppenebene rechnen)
 - Disaggregation der Daten
(„Complete pooling“)
 - Gruppen separat analysieren
(„no pooling“)
- Achtung: Alle drei „work-arounds“ führen zu verzerrten Ergebnissen!

- Reduktion der Anzahl der Personen-Werte auf Anzahl der Gruppierungsvariablen-Fälle.
 - Statt „1200 Schüler*innen in 47 Klassen“ schaut man sich also nur noch „47 Klassen an“
- Aggregation der Personen-Werte in den Prädiktoren und dem Kriterium durch Mittelwertbildung innerhalb der Gruppierungseinheiten.
- Dann Regressionsschätzung der mittleren Kriteriumswerte aus mittleren Prädiktor-Werten durchführen.

Ausgangsdatensatz:

*Unbenannt2 [DatenSet1] - SPSS Daten-Editor

	id	Therapeuten_id	Neurotizismus_Score	Symptomscore
1	1	1	66	34
2	2	1	34	23
3	3	1	65	44
4	4	1	43	25
5	5	1	68	43
6	6	2	34	23
7	7	2	23	54
8	8	2	56	23
9	9	2	32	55
10	10	2	54	23
11	11	3	65	47
12	12	3	34	48
13	13	3	55	55
14	14	3	33	60
15	15	3	44	34

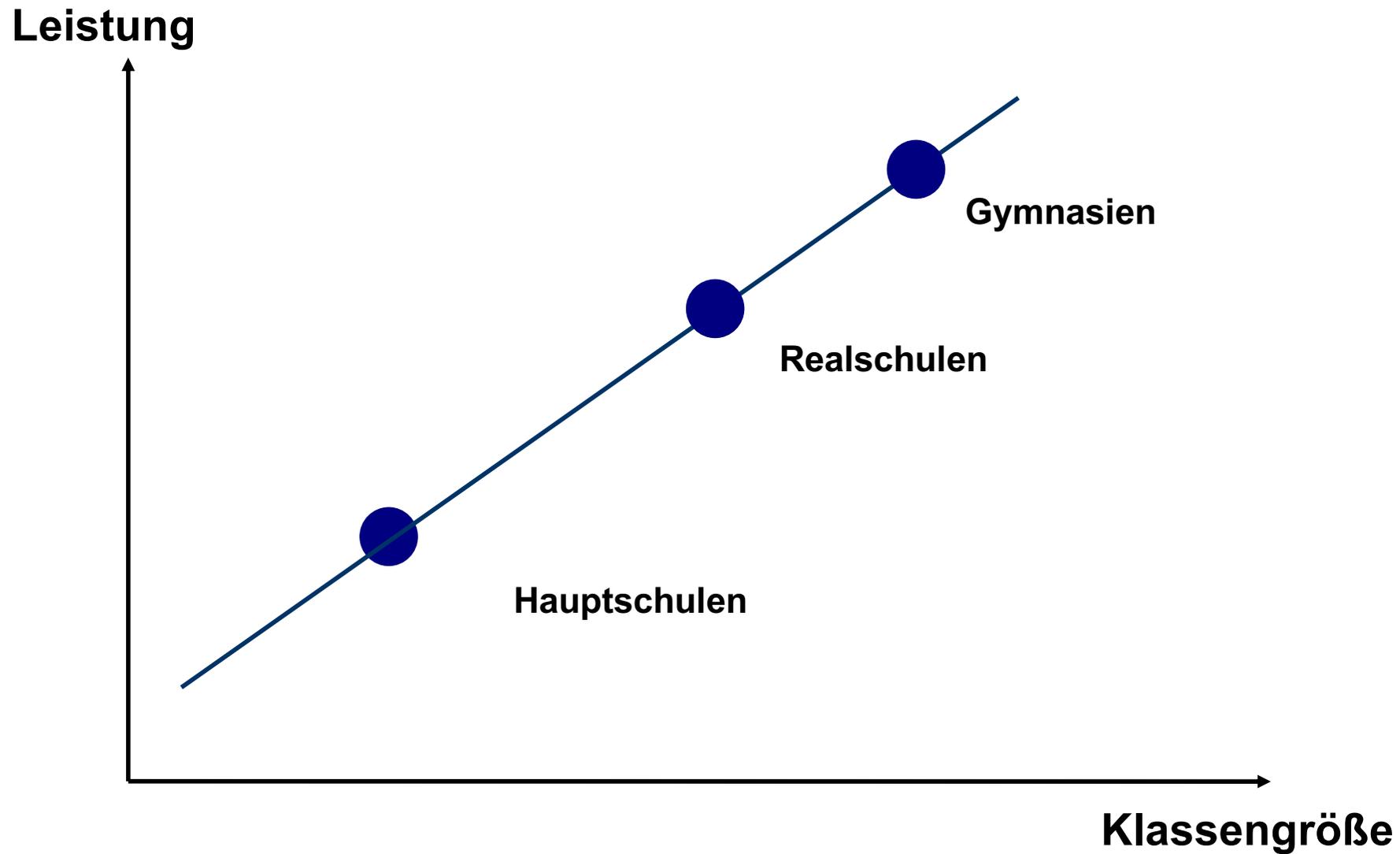


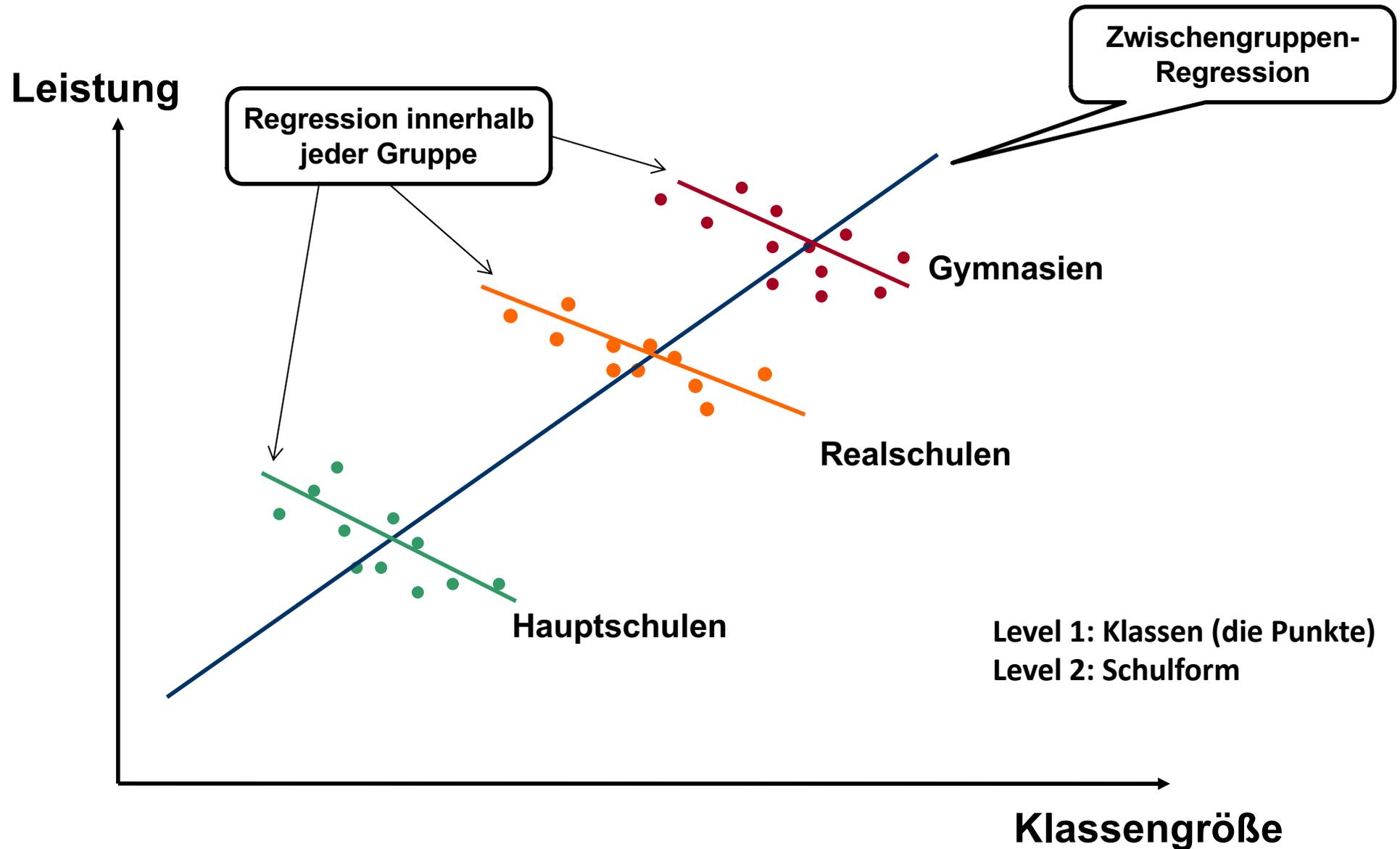
Aggregierter Datensatz:

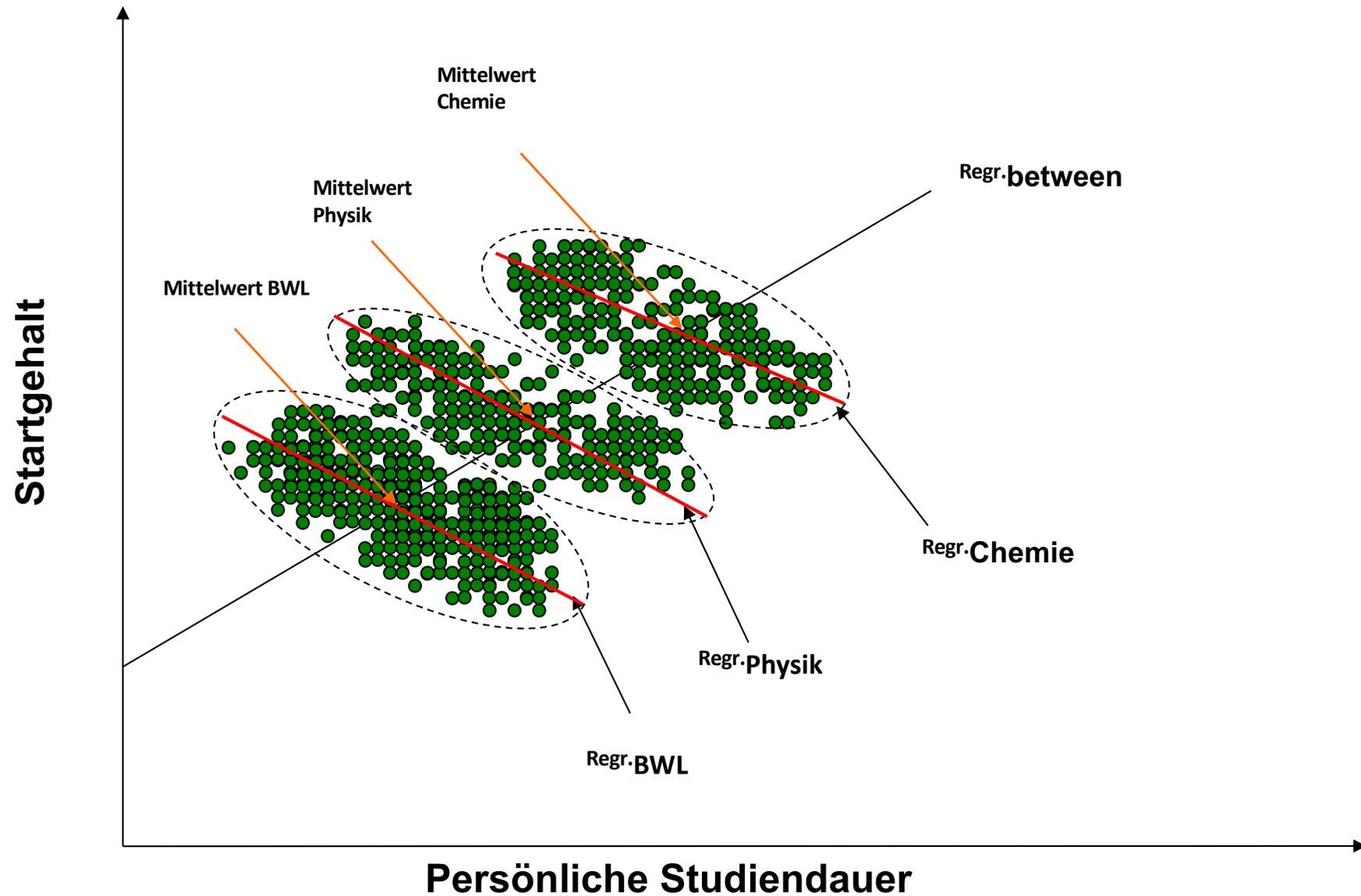
*Unbenannt3 [DatenSet2] - SPSS Daten-Editor

	Therapeutengruppe	Mittelwert_Neurot	Mittelwe
1	1	55	
2	2	40	
3	3	46	

- **„Shift of meaning“**: Die Interpretation der Koeffizienten bezieht sich nur noch auf die Gruppen-Ebene.
 - Bsp.: Mittelwert der Beurteilung von Arbeitsbedingungen als Index des Organisationsklimas bezieht sich nur auf die Firma, nicht auf die Angestellten.
- **Ökologischer Fehlschluss**: Korrelation auf Gruppen-Ebene kann keine Aussagen auf Individual-Ebene machen.
- Keine Analyse von Interaktionen zwischen Gruppierungs- und Individualebene möglich:
 - Bsp.: Gibt es eine Interaktion zwischen Begabung und Unterrichtsform (Gesamtklassen-Unterricht vs. Leistungsgruppen-Unterricht) hinsichtlich Schulleistung ?







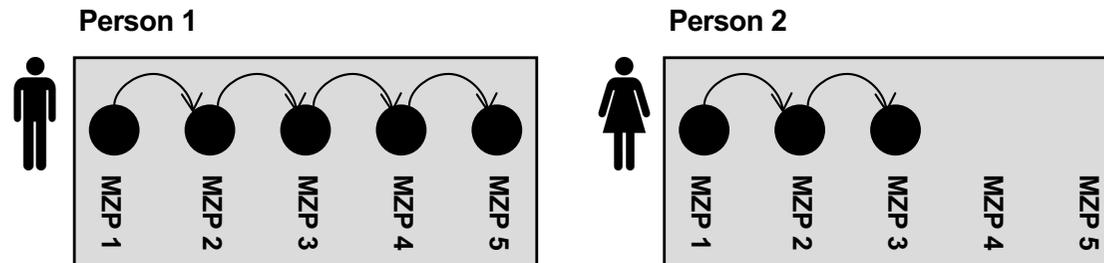
- Speed-accuracy-trade off
 - Höhere Geschwindigkeit geht mit **weniger** Fehlern einher (zwischen Personen)
 - Höhere Geschwindigkeit geht mit **mehr** Fehlern einher (innerhalb von Personen)
- Training und Müdigkeit
 - Mehr Training führt zu **weniger** Müdigkeit (zwischen Personen)
 - Mehr Training führt zu **mehr** Müdigkeit (innerhalb von Personen)

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychol. Rev.* 89, 483–506.

Puetz, T. W., O'Connor, P. J., & Dishman, R. K. (2006). Effects of chronic exercise on feelings of energy and fatigue: A quantitative synthesis. *Psychological Bulletin*, 132(6), 866–876. doi:10.1037/0033-2909.132.6.866

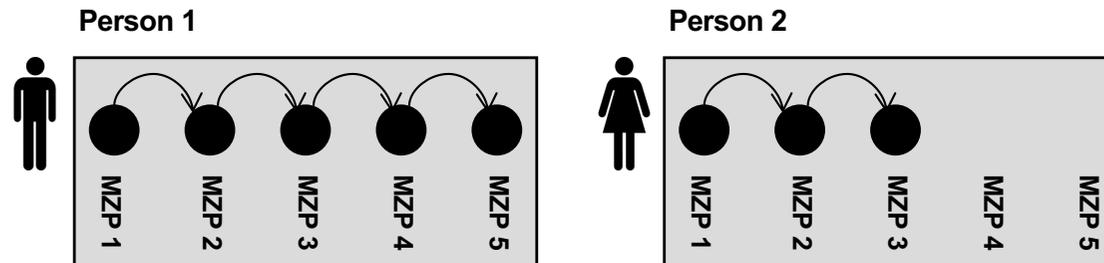
For a summary, see Kievit, R. A., & Epskamp, S. (2013). Simpson's paradox in psychological science: a practical guide, 1–14. doi:10.3389/fpsyg.2013.00513/abstract

- Man kann prinzipiell nicht von Zusammenhängen einer Aggregationsebene auf die Zusammenhänge auf einer anderen Aggregationsebene schließen!
- Vgl. Persönlichkeitspsychologie:
 - Korrelatives interindividuelles Design zeigt:
Personen, die vegan leben, sind glücklicher
(Zusammenhang zwischen Personen / interindividuelle Variabilität)
 - Schluss: Wenn ich anfangs, vegan zu essen, dann werde ich glücklicher!
(Zusammenhang innerhalb einer Person / intraindividuelle Variabilität)
 - Stop! Diesen Schluss darf man nicht automatisch ziehen.



Der Schluss von einer Ebene auf die andere kann nur unter zwei Bedingungen – den sogenannten *Ergodizitätsannahmen* – gezogen werden

1. **Homogenität:** Das selbe statistische Modell gilt für jede Gruppierungseinheit auf L2 (z.B. Personen)
 - Gegenbeispiel: Big 5-Struktur. Man kann intraindividuell mehrfach messen und eine Faktorenstruktur für jede einzelne Person berechnen (d.h., wie sind Persönlichkeits-States intraindividuell organisiert?)
 - Die individuellen State-Modelle sind sehr unterschiedlich – manche Personen hatten 2 Faktoren, manche 3, manche 4, ...

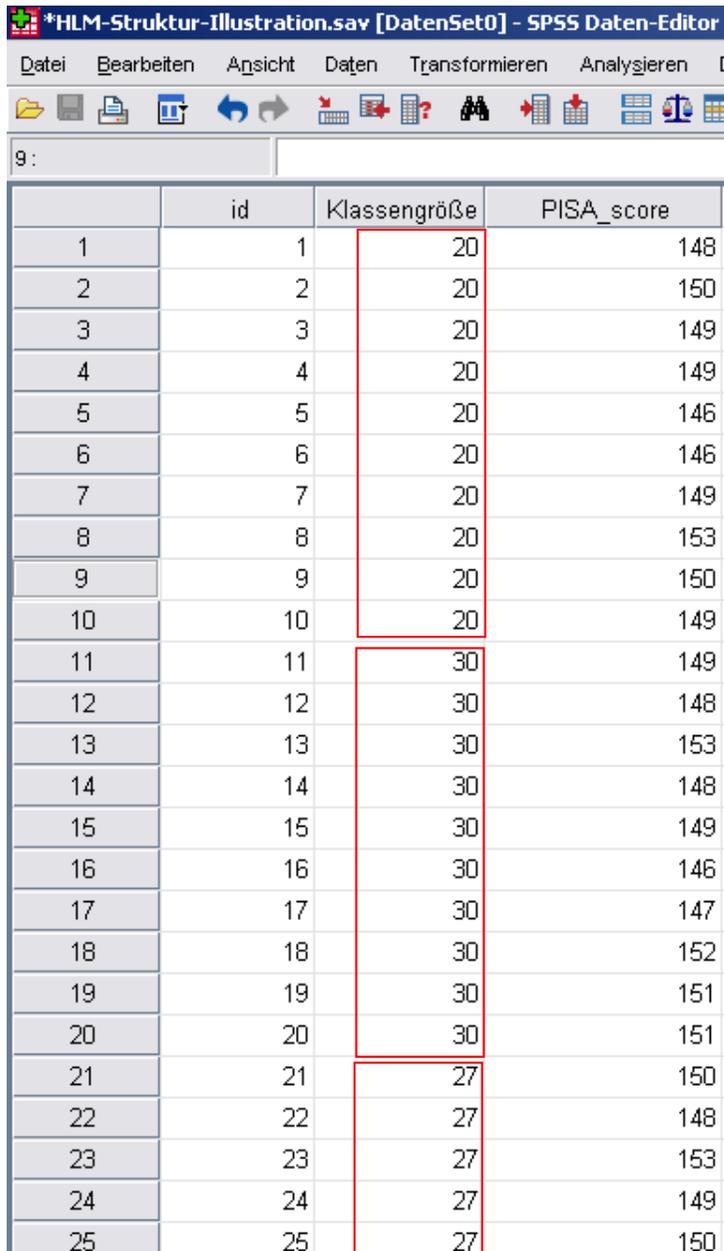


2. **Stationarität:** Der psychologische Prozess auf L1 muss stationär sein (d.h., konstante statistische Charakteristika über die Zeit aufweisen – gleicher MW, SD, Faktorladungen, etc. an jedem Zeitpunkt)
- Gegenbeispiele: Lernen, Habituation, Entwicklungsprozesse, ... (nahezu jeder beliebige psychologische Prozess)
 - Fazit: Die Ergodizitätsbedingungen sind so strikt und typischerweise unplausibel, dass man generell *nicht* von einer Ebene auf die andere schließen kann.

- Wie könnte man nun mit herkömmlichen regressionsanalytischen Methoden gruppierte Daten analysieren?
 - Aggregation der Daten
(d.h. nur auf Gruppenebene rechnen)
 - **Disaggregation der Daten**
(„Complete pooling“)
 - Gruppen separat analysieren
(„no pooling“)
- Alle drei „work-arounds“ führen zu falschen Ergebnissen!

Hauptproblem: Ökologischer
Fehlschluss

- So tun, als gäbe es keine Gruppierungsstruktur
 - „Vervielfachung“ der L2-Variablen
 - Jeder Person wird in einer neuen Variable der jeweilige Wert der Gruppierungs-Variablen zugewiesen.
 - Jeder Schülerin die Größe ihrer Klasse
 - Jedem Patienten der Code der Klinik, in der er behandelt wird
 - Jeder Angestellten der Durchschnittsverdienst einer Abteilung
- Dann Berechnung einer Regression mit dieser neuen Variable als Prädiktor.



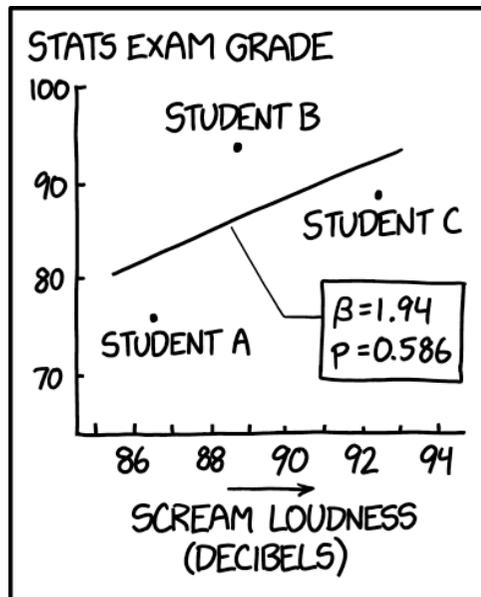
The screenshot shows the SPSS Daten-Editor interface. The title bar reads '*HLM-Struktur-Illustration.sav [DatenSet0] - SPSS Daten-Editor'. The menu bar includes 'Datei', 'Bearbeiten', 'Ansicht', 'Daten', 'Transformieren', and 'Analysieren'. The toolbar contains various icons for file operations and data manipulation. The main window displays a data table with the following columns: 'id', 'Klassengröße', and 'PISA_score'. The data is as follows:

	id	Klassengröße	PISA_score
1	1	20	148
2	2	20	150
3	3	20	149
4	4	20	149
5	5	20	146
6	6	20	146
7	7	20	149
8	8	20	153
9	9	20	150
10	10	20	149
11	11	30	149
12	12	30	148
13	13	30	153
14	14	30	148
15	15	30	149
16	16	30	146
17	17	30	147
18	18	30	152
19	19	30	151
20	20	30	151
21	21	27	150
22	22	27	148
23	23	27	153
24	24	27	149
25	25	27	150

- Jeder Schüler*in wird der jeweilige Wert der Klassengröße zugeordnet.
- Dann wird der PISA-Score aus der Variable „Klassengröße“ über eine einfache Regression geschätzt.
- Aus ursprünglich drei unabhängigen Werten (drei Gruppen) werden $3 \times 10 = 30$ Werte gebildet, die wie 30 unabhängige Werte behandelt werden.

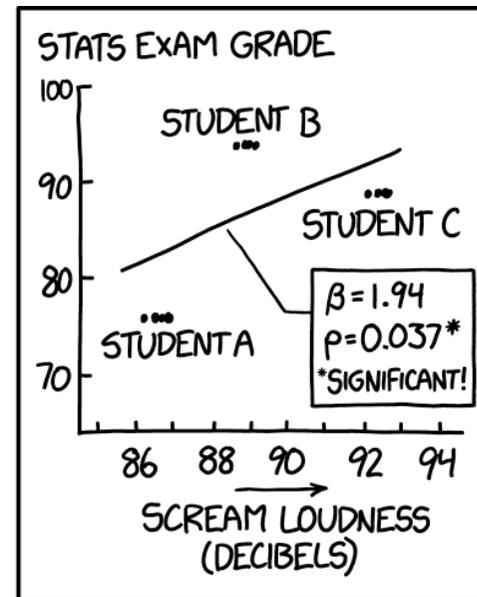
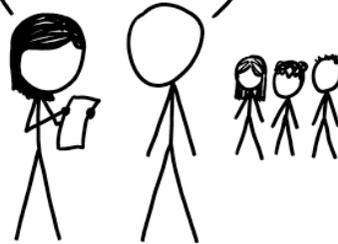
Folge:

- a) Künstliche Vervielfachung der Stichprobengröße auf Gruppenebene.
- b) Behandlung der eigentlich abhängigen Werte als *unabhängige* Werte.



DARN, NOT SIGNIFICANT.

WE NEED MORE DATA.
HAVE THEM EACH TRY
YELLING INTO THE MIC
A FEW MORE TIMES.



PERFECT!

ARE YOU SURE
WE'RE DOING
SLOPE HYPOTHESIS
TESTING RIGHT?

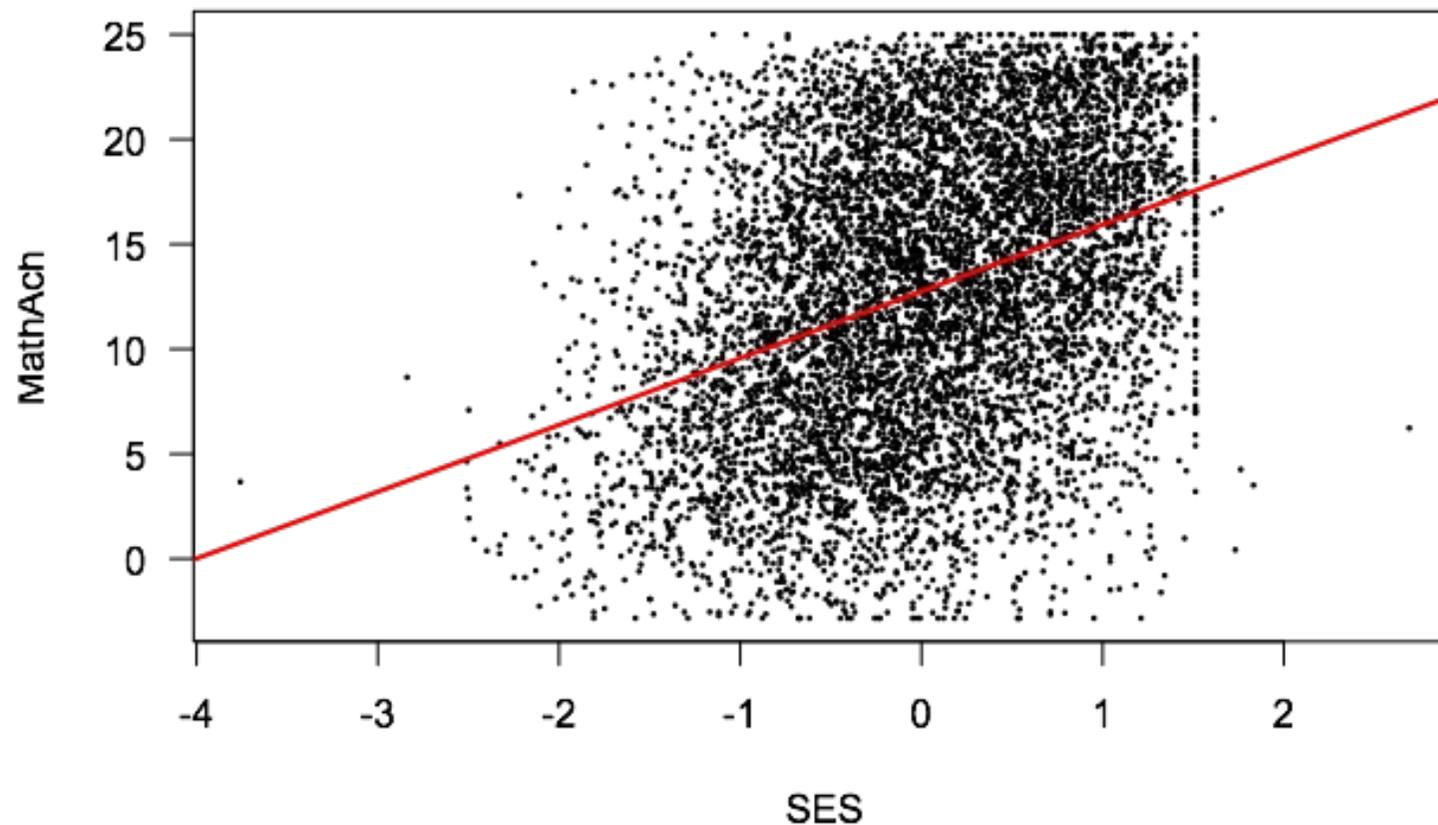


Complete pooling / Disaggregation

```
lm(MathAch ~ SES, HSB)
```

$b_0 = 12.75*** [12.60; 12.90]$

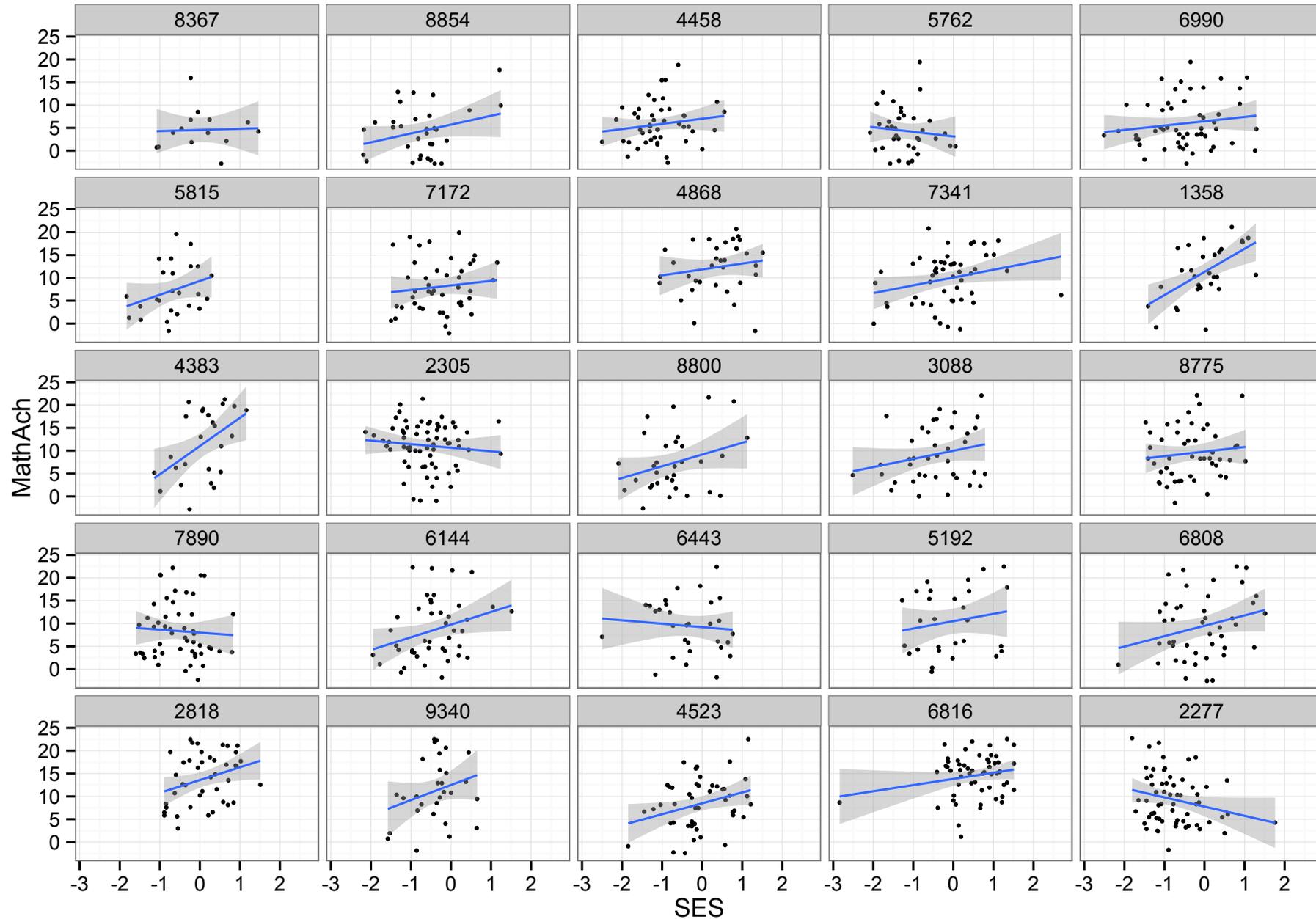
$b_1 = 3.18*** [2.99; 3.37]$

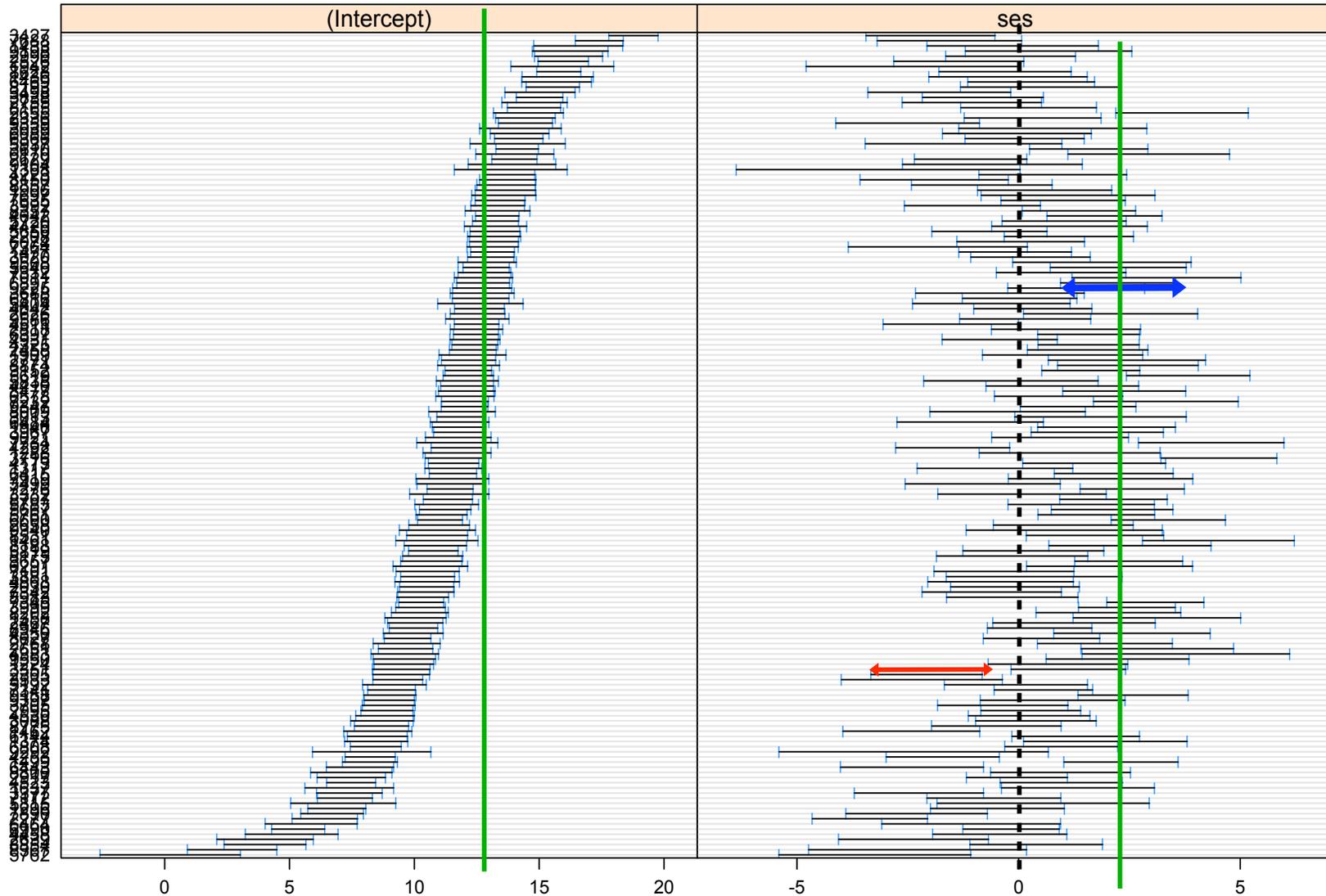


- Probleme bei Disaggregation:
 - Werden Unterschiede zwischen Gruppen untersucht (bspw. Leistungsunterschiede zwischen Klassen unterschiedlicher Größe):
Alpha-Fehler-Kumulierung bei L2-Prädiktoren wegen **künstlicher** Vervielfachung des Stichprobenumfangs
→ Power von Hypothesentests wird überschätzt
- *Ignoriert* Unterschiede zwischen Gruppen
 - Genau eine Regressionlinie für alle

- Wie könnte man nun mit herkömmlichen regressionsanalytischen Methoden gruppierte Daten analysieren?
 - Aggregation der Daten
(d.h. nur auf Gruppenebene rechnen)
 - Hauptproblem: Ökologischer Fehlschluss
 - Disaggregation der Daten
(„Complete pooling“)
 - Hauptproblem: Erhöhte Typ-I-Fehlerrate bei L2-Variablen; ignoriert Variabilität der Gruppen
 - **Gruppen separat analysieren**
(„no pooling“)
- Alle drei „work-arounds“ führen zu falschen Ergebnissen!

Plot für 25 von 160 Schulen





mean $b_0 = 12.60$

mean $b_1 = 2.20$ $\frac{4}{0}$

Complete

pooling: $b_0 = 12.75^{***} [12.60; 12.90]$

$b_1 = 3.18^{***} [2.99; 3.37]$

Gruppen separat analysieren („no pooling“)

- Jede Gruppe bekommt ihre eigene Regressionslinie
- Bei z.B. 160 Klassen bekommt man dann:
 - 160 Intercepts
 - 160 Slopes
- *Überschätzt* Unterschiede zwischen Gruppen:
 - Einzelne Gruppen mit wenigen Datenpunkten/extremen Slopes
 - → es ist plausibel anzunehmen, dass dies Ausreißer sind; man sollte das Wissen über die anderen 159 Klassen nutzen, um die extreme Regression etwas zu korrigieren (später mehr dazu).
 - Overfitting der Daten in jeder einzelnen Gruppe
- Bessere Variante: Diese Modellparameter auf Gruppenebene *gewichtet* aggregieren.

- Wie könnte man nun mit herkömmlichen regressionsanalytischen Methoden gruppierte Daten analysieren?

➤ Aggregation der Daten
(d.h. nur auf Gruppenebene rechnen)

Hauptproblem: Ökologischer
Fehlschluss

➤ Disaggregation der Daten
(„Complete pooling“)

Hauptproblem: Erhöhte Typ-I-
Fehlerrate bei L2-Variablen;
ignoriert Variabilität der Gruppen

➤ **Gruppen separat analysieren**
(„no pooling“)

Hauptproblem: Overfitting in
einzelnen Gruppen; wie fasst
man die vielen Modellparameter
optimal zusammen?

- Alle drei „work-arounds“ führen zu falschen Ergebnissen!

- **Fazit** (Cohen, 2003): „*Clustered data historically have been the cause of great statistical hand wringing and analytic acrobatics.*“
 - Man braucht also einen Regressionsansatz, der die Beziehungen der Gruppierungs- und Individualebene **simultan**, d.h. in einer einzigen Gleichung, berücksichtigt und die Einflüsse von Prädiktoren jeder Ebene schätzen kann.
- Lösung:
- Multilevel Modelle**, bzw. spezifischer, **Lineare gemischte Modelle**

- Gruppierte /geclusterte Datenstrukturen verletzen eine zentrale Annahme der normalen multiplen Regression: die Unabhängigkeit der Fehlerterme
- Wie kann man diese (Un-)abhängig feststellen?
 - Graphisch die Residuen pro L2-Einheit plotten
 - Intraklassenkoeffizient berechnen
- Drei scheinbare „Lösungen“ → **nicht verwenden!**
 - *Aggregation: Regression nur auf Gruppenebene rechnen.*
Problem: ökologischer Fehlschluss
 - *Disaggregation: Gruppenstruktur ignorieren („complete pooling“).*
Problem: Erhöhter Typ-I-Fehler für L2-Variablen;
Varianz zwischen Gruppen ignoriert.
 - *Jede Gruppe einzeln rechnen („no pooling“).* Problem: Overfitting; wie fasst man die einzelnen Modelle optimal zusammen?