TECHNISCHE UNIVERSITÄT
IN DER KULTURHAUPTSTADT EUROPAS
CHEMNITZ

Professur Psychologie digitaler Lernmedien

Institut für Medienforschung Philosophische Fakultät

Einführung in die Statistik

Mehrebenenmodelle

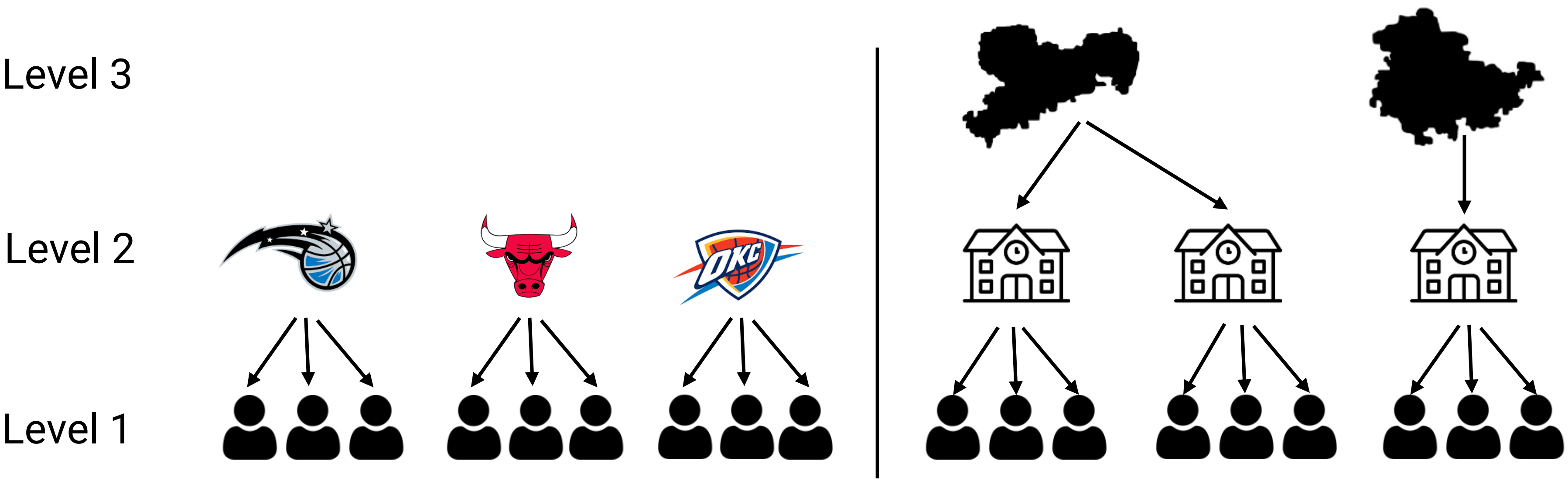
Prof. Dr. Günter Daniel Rey

Überblick

- Hierarchische Daten
- Spezifika hierarchischer Datenstrukturen
- Feste und zufällige Effekte in Mehrebenenmodellen
- Vorgehen bei der Mehrebenenmodellierung
- Modellgüte bei Mehrebenenmodellen
- Voraussetzungen der Mehrebenenmodellierung
- Zentrierung bei Mehrebenenmodellen

Hierarchische Daten
(z. B. Rudolf & Vogel-Blaschka, 2023)

- empirische Daten sind in vielen Fällen nicht unabhängig voneinander, sondern weisen hierarchische Strukturen auf
- fehlende Unabhängigkeit der Daten verletzt die Voraussetzung zahlreicher Verfahren des Allgemeinen Linearen Modell
- Beispiele:

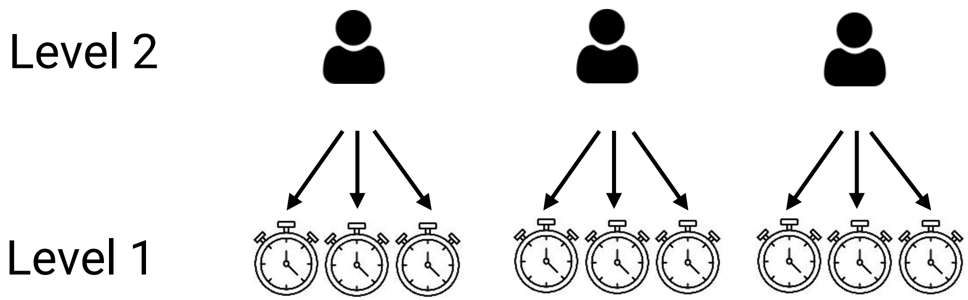


Hierarchische Daten
(z. B. Rudolf & Vogel-Blaschka, 2023)

- Warum können Daten aus hierarchische Strukturen nicht als unabhängig voneinander betrachtet werden?
- Beispiel: Zufriedenheit von Schülerinnen und Schülern mit der Unterrichtsqualität (Kriteriumsvariable)
  - Messung erfolgt auf Level 1 (Individuen), bei denen durch Zugehörigkeit zur gleichen übergeordneten Strukturen eine höhere Ähnlichkeit vorliegt als bei Zugehörigkeit zu verschiedenen Strukturen
  - Schüler/-innen in der gleichen Klasse (Level 2) sind sich ähnlicher als Schüler/-innen unterschiedlicher Klassen (gleicher Stundenplan, gleiche Lehrer etc.)
  - Schüler/-innen der gleichen Schule (Level 3) sind sich ähnlicher als Schüler/-innen unterschiedlicher Schulen (gleiches Gebäude, gleiche Lage, gleiche Ausstattung)
  - Schüler/-innen im gleichen Bundesland (Level 4) sind sich ähnlicher als Schüler/-innen unterschiedlicher Bundesländer (gleiche Gesetze, Bildungsziele, Regelungen etc.)

Hierarchische Daten
(z. B. Rudolf & Vogel-Blaschka, 2023)

- Abhängigkeit von Messwerten liegt ebenfalls vor bei mehrfachen Messungen bei den gleichen Versuchspersonen (within-subject)
  - Achtung: erst sinnvoll ab drei Messzeitpunkten
- einzelne Messungen werden von Persönlichkeitsmerkmalen des Individuums beeinflusst → Daten derselben Person sind untereinander ähnlicher als Daten unterschiedlicher Personen
- es sind auch gemischte Designs möglich, z. B. Level 1: Messzeitpunkte, Level 2: Studierende, Level 3: Universitäten



Hierarchische Daten
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 124)

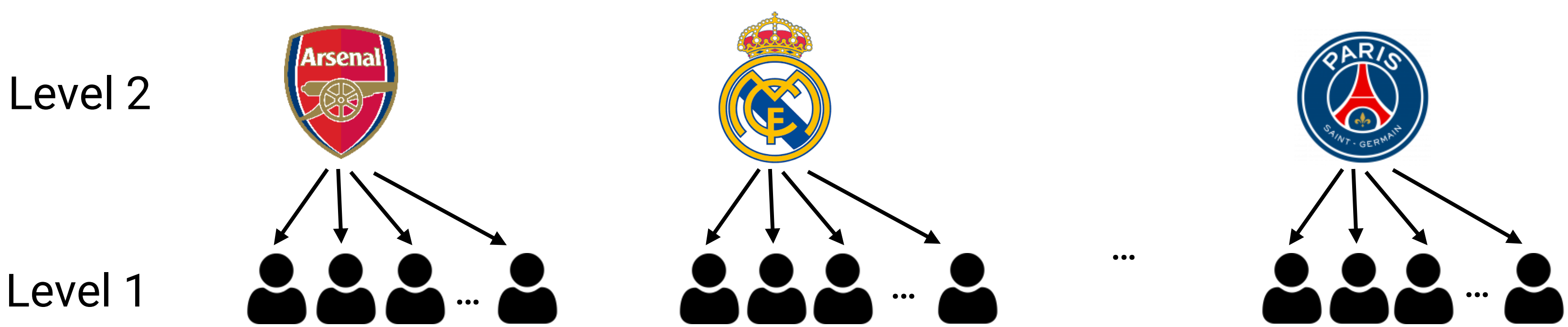
Typische Erkenntnisinteressen bei Mehrebenenmodellen

- Berechnung des Anteils der Varianz in der Kriteriumsvariablen, der auf Unterschiede der Level-2-Einheiten zurückzuführen ist
- Untersuchung des Zusammenhangs von Prädiktoren und Kriterium auf Level 1 unter Berücksichtigung der Mehrebenenstruktur
- Beurteilung des Einflusses von Level-1-Prädiktoren und von Level-2-Prädiktoren auf die Kriteriumsvariable
- Untersuchung von Moderatoreffekten von Level-2-Prädiktoren auf Zusammenhänge von Level-1-Prädiktoren und Kriterium
- Beurteilung von Varianzaufklärungen auf Level 1 und auf Level 2

Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023)

Beispiel für die nachfolgenden Erklärungen:

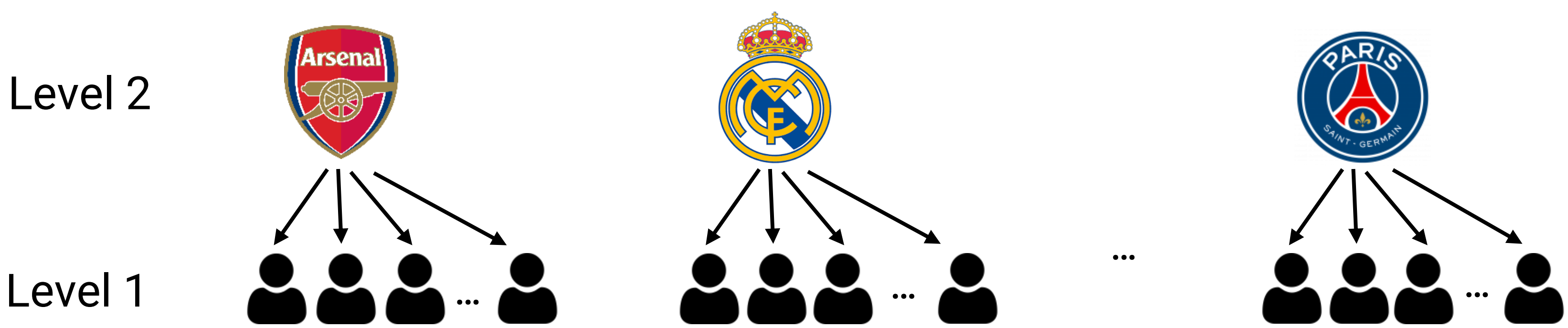
- Einfluss von Stress (Prädiktor) auf das Leistungsvermögen (Kriterium) von Profi-Fußballerinnen bzw. -fußballern (Level 1) in Abhängigkeit von ihrem Klub (Level 2), da unterschiedliche Effekte der Klubzugehörigkeit



Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

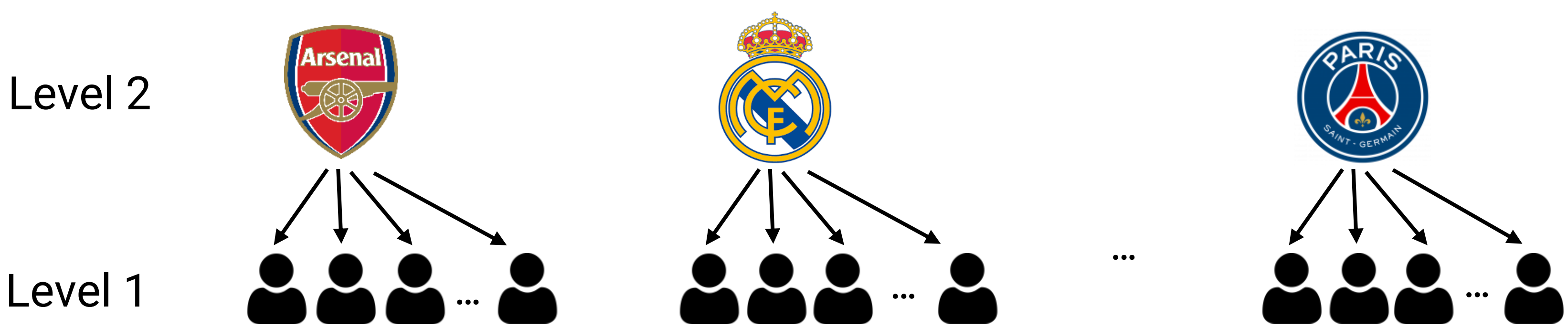
- Daten werden auf mehreren hierarchisch geordneten Ebenen erhoben.
    - Im Beispiel werden zwei Ebenen (Level) in die Betrachtung einbezogen: Level 1 beinhalten die einzelnen Spieler/-innen, Level 2 stellt die Ebene der Fußballclubs dar.



Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

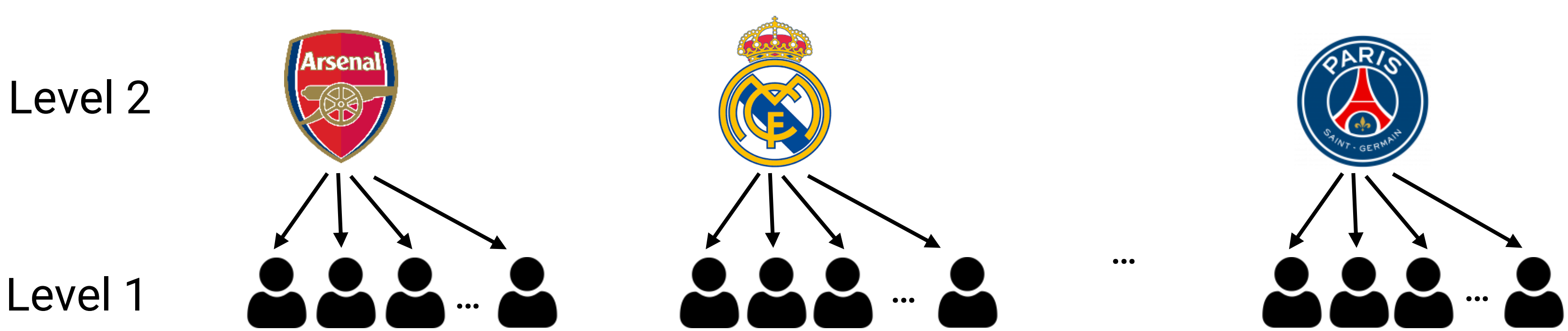
- Die Einheiten innerhalb aller Ebenen sind klar definiert und beobachtbar.
  - Jede/r einzelne Sportler/-in und jeder einzelne Fußballclub ist klar bestimmt und sowohl von den Sportler/-innen und den Fußballclubs können Daten erhoben werden.



Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

- Jede Einheit einer niedrigeren Ebene ist eindeutig einer Einheit der übergeordneten Ebene zugeordnet.
  - Jede/r einzelne Sportler/-in spielt zum Zeitpunkt der Datenerhebung bei genau einem Verein, die Zuordnung ist somit eindeutig.



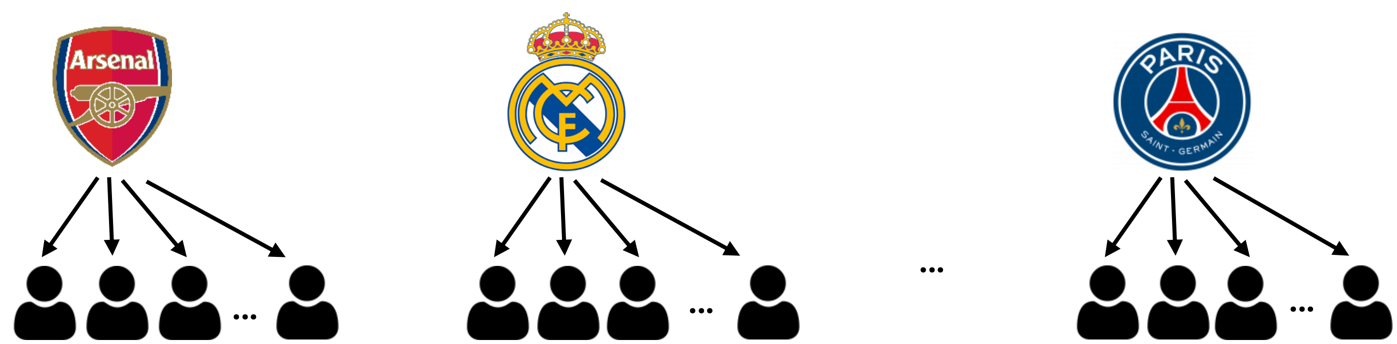
Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

- Die Gruppen der unterschiedlichen Ebenen können verschieden groß sein.
  - Das Ergebnis der Untersuchung ist nicht davon abhängig, ob jeder Verein genau gleich viele Spieler/-innen hat.

Level 2

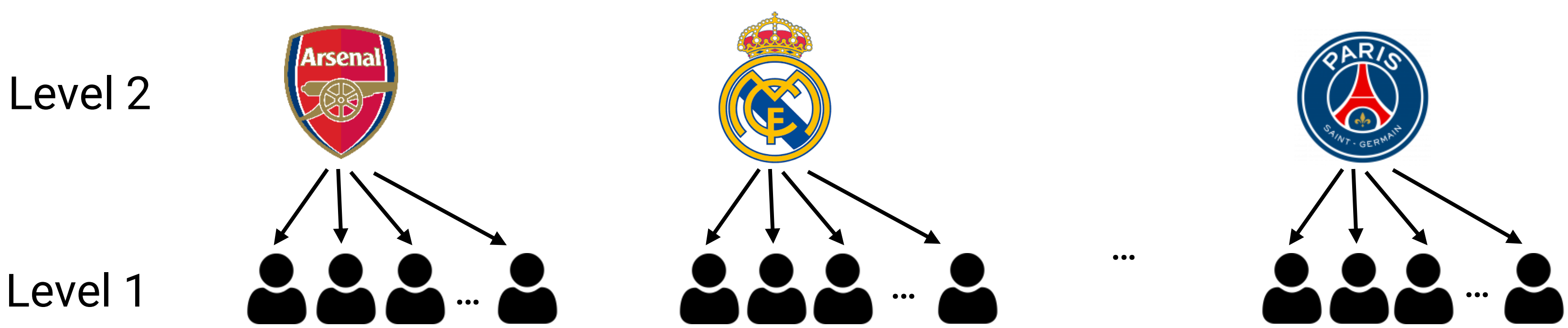
Level 1



Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

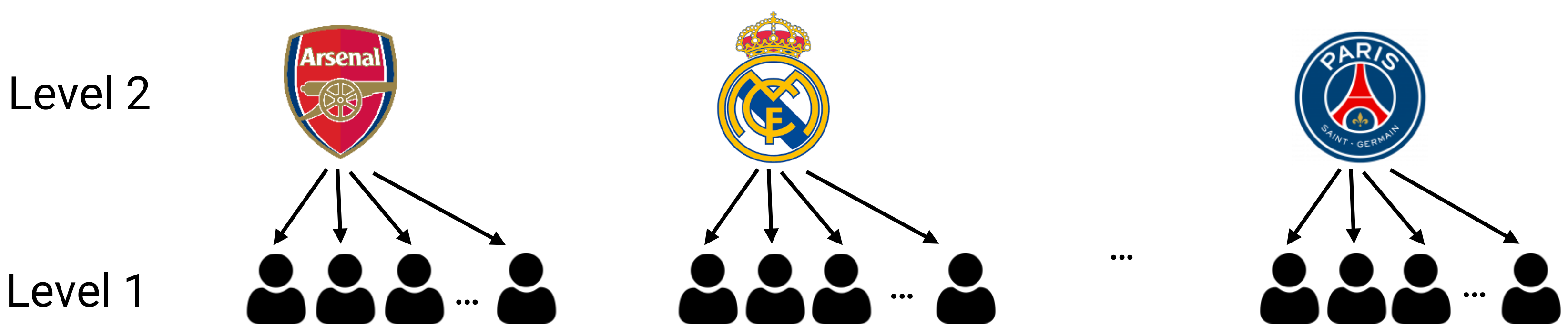
- Das Kriterium wird jeweils auf der untersten Ebene (Level 1) gemessen.
  - Das Leistungsvermögen der Spieler/-innen wird bei jeder einzelnen Person gemessen. Eigenschaften des Clubs (Level 2) können in den hier betrachteten Modellen lediglich als Prädiktor wirken, nicht aber als Kriterien.



Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

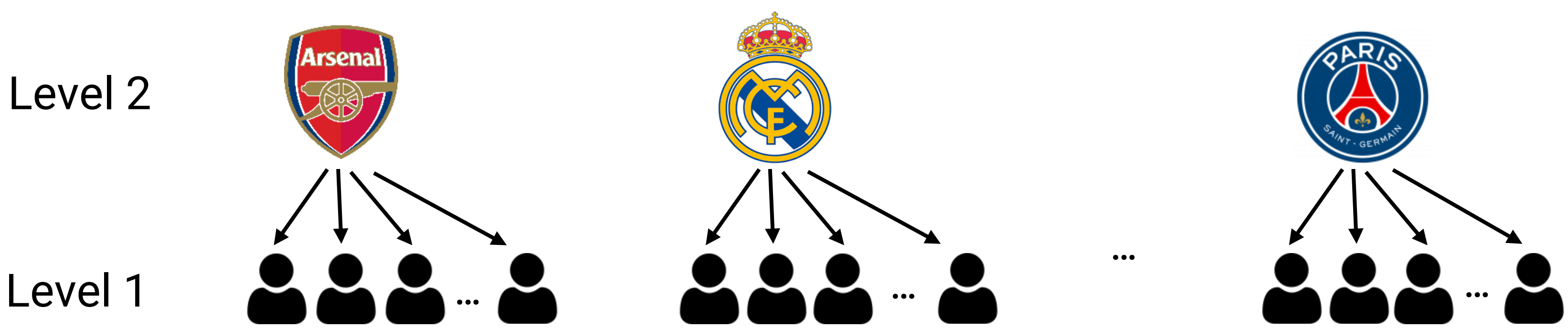
- Prädiktoren können auf allen Ebenen erhoben werden.
  - Im Gegensatz zum Kriterium können Prädiktoren sowohl bei den Spieler/-innen (Level 1) als auch bei den Clubs (Level 2) erhoben und in das Modell aufgenommen werden. Im Beispiel könnte das Leistungsvermögen von individuellen Merkmalen (Motivation, Selbstwirksamkeit, Zielorientierung etc.) oder Merkmalen des Clubs (mentale Betreuung, Trainer, Saisonziele etc.) abhängig sein.



Spezifika hierarchischer Datenstrukturen
(z. B. Rudolf & Vogel-Blaschka, 2023, S. 126)

Besonderheiten hierarchischer Datenstrukturen

- Beobachtung in Einheiten übergeordneter Ebenen sind in der Regel nicht unabhängig voneinander. Sie ähneln sich stärker als Beobachtungen aus verschiedenen übergeordneten Einheiten.
  - Unterschiedliche Rahmenbedingungen in unterschiedlichen Clubs sorgen dafür, dass sich Spieler/-innen typischerweise ähnlicher sind, wenn sie im gleichen Verein sind als wenn sie in unterschiedlichen Vereinen sind.



Spezifika hierarchischer Datenstrukturen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127)

Warum greift eine reguläre lineare Regressionsanalyse zu kurz?

- Klassische Regression kann jeweils nur „eine Ebene“ berücksichtigen, d. h. in unserem Beispiel könnte man die Daten aller Sportler pro Verein aggregieren und mitteln → Auswertung auf Level der Vereine
  - Individuelle Unterschiede zwischen den Fußballern auf Level 1 werden ignoriert ➔ einerseits resultiert dieses Vorgehen in einer reduzierten Teststärke; andererseits ist eine sachgerechte Interpretation der Ergebnisse nicht möglich
- Alternative: regressionsanalytische Auswertung auf Basis der Fußballer (Level 1) ohne Rücksicht der übergeordneten Vereinsebene (Level 2)
  - Missachtung der Voraussetzung unabhängiger Messfehler der Fußballer → starke Erhöhung des Fehlers erster Art
  - erhöhte Wahrscheinlichkeit inhaltlicher Fehlschlüsse, wenn Daten der Gruppenebene unberücksichtigt bleiben und nur individuelle Fälle betrachtet werden → Simpson-Paradox (Simpson, 1951)

Spezifika hierarchischer Datenstrukturen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127)

Zusammenhang zwischen Verantwortung im Job und Arbeitszufriedenheit bei Unternehmen

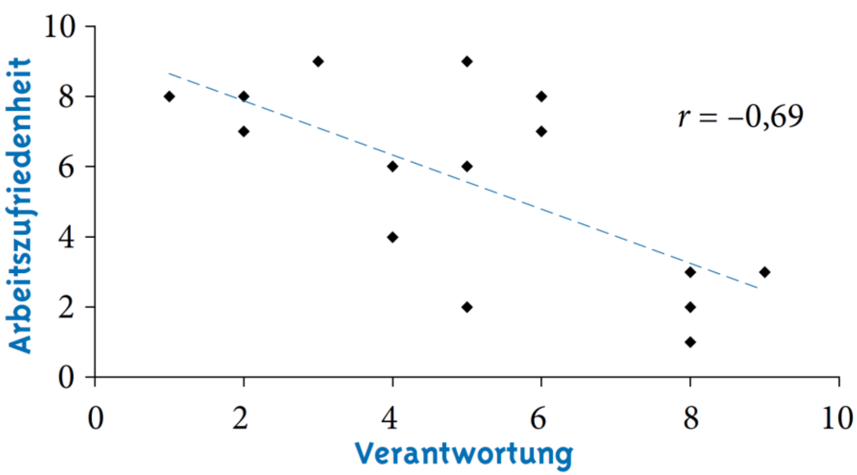


Abbildung 20.1 Streudiagramm für das Datenbeispiel aus Tabelle 20.1: Regression über alle Messwerte hinweg

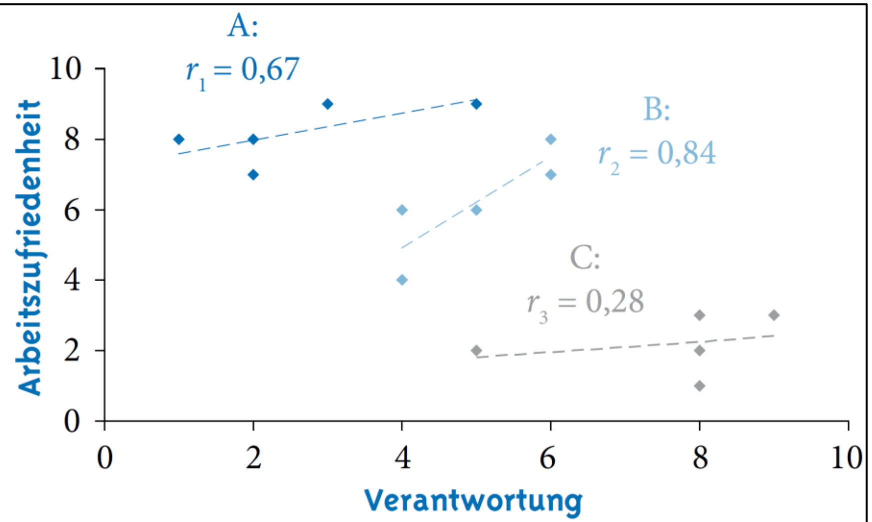


Abbildung 20.2 Streudiagramm für das Datenbeispiel aus Tabelle 20.1: Regression für jede Firma getrennt

Quelle: Eid, Gollwitzer und Schmitt (2017)

Negative Korrelation von $r = -.69$
→ widerspricht gängigen Modellen und Befunden, warum?

Daten entstammen drei Abteilungen einer Firma, die jeweils die erwartete positiven Korrelationen aufweisen

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127)

Feste vs. zufällige Effekte

- Feste Effekte (Fixed Effects) = "Gleichschaltung"
  → Wir zwingen das Modell dazu, für alle Gruppen denselben Wert zu berechnen. Das Modell drückt aus: „Es gibt nur eine Steigung und einen Startwert für alle Fußballer/-innen.“ Unterschiede zwischen den Clubs werden ignoriert oder als reiner "Messfehler" interpretiert.
- Zufällige Effekte (Random Effects) = "Individualität"
  → Wir geben dem Modell die Erlaubnis, dass jede Gruppe (jeder Club) seinen eigenen Wert haben darf.
  - Random Intercept: Wir erlauben jedem Club, ein eigenes Startniveau an Leistungsvermögen zu haben.
  - Random Slope: Wir erlauben jedem Club, einen eigenen Zusammenhang (Steigung) zwischen Stress und Leistungsvermögen zu haben.

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127ff)

Ausgangspunkt für weitere Betrachtungen: bekanntes Modell der einfachen linearen Regression

$$y_i = b_0 + b_1 \cdot x_i + r_i \ (i = 1, \ldots, n)$$

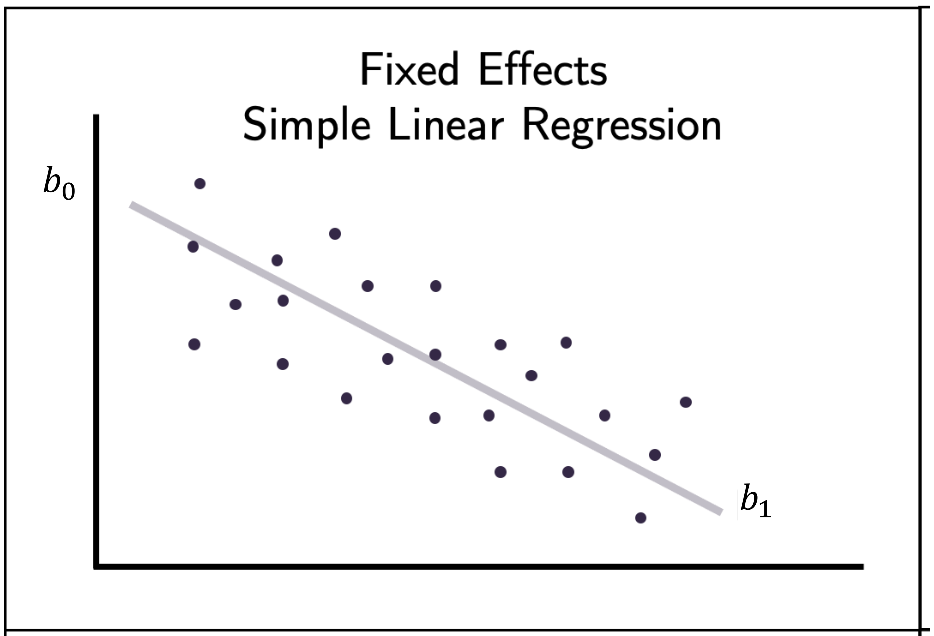
y_i: Wert der Kriteriumsvariablen y des i-ten Spielers
x_1: Wert der Prädiktorvariablen x des i-ten Spielers
r_i: Residuum des i-ten Spielers
b_0: Regressionskonstante (Intercept)
b_1:Regressionssteigung (Slope)
n: Anzahl der Spieler

- im klassischen Modell gibt zwei fixen Werte, der das Mittel für alle Fälle darstellt
- Intercept b_0:
  - erwarteter Wert des Leistungsvermögens, wenn der Stress 0 ist (bzw. bei Zentrierung: erwarteter Wert bei durchschnittlichem Stressempfinden
- Slope (b_1):
  - mittlerer Zusammenhang zwischen Stress und Leistungsvermögen über alle Spieler/-innen hinweg (ohne Berücksichtigung des Clubs)

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127ff)

Ausgangspunkt für weitere Betrachtungen: bekanntes Modell der einfachen linearen Regression

$$y_i = b_0 + b_1 \cdot x_i + r_i \ (i = 1, \dots, n)$$



negative Steigung zeigt, dass im Durchschnitt das Leistungsvermögen (Kriterium) bei zunehmendem Stress (Prädiktor) sinkt

ABER: hier wird Level 2 (Vereine) nicht berücksichtigt

Quelle: Midway (o.D.), Figure 9.1 Link

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127ff)

Berücksichtigung unterschiedlicher Regressionskonstanten in Abhängigkeit des Vereins (random intercept, fixed slope)

$$y_{ij} = b_{0j} + b_1 \cdot x_{ij} + r_{ij} \; (i = 1, \ldots, n_j; j = 1, \ldots, m) \quad \text{mit} \quad b_{0j} = b_0 + u_{0j}$$

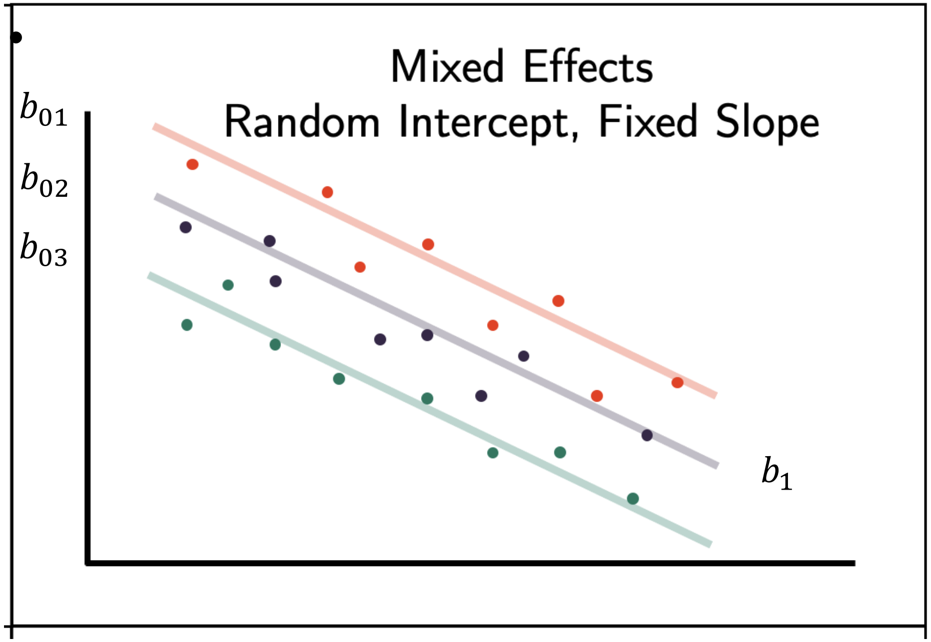
y_{ij}: Wert der Kriteriumsvariablen y des i-ten Spielers im j-ten Verein
x_{ij}: Wert der Prädiktorvariablen x des i-ten Spielers im j-ten Verein
r_{ij}: Residuum des i-ten Spielers im j-ten Verein
b_{0j}: Regressionskonstante (Intercept) für den j-ten Verein
u_{0j}: Abweichung der Regressionskonstante des j-ten Vereins vom mittleren Intercept
b_1: Regressionssteigung (Slope)
n_j: Anzahl der Spieler im j-ten Verein
m: Anzahl der Vereine

- Intercept b_{0j}:
  - parallele Geraden, die an unterschiedlichen Stellen die y-Achse schneiden; Zusammenhang zwischen Prädiktor und Kriterium ist bei allen gleich

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127ff)

Berücksichtigung unterschiedlicher Regressionskonstanten in Abhängigkeit des Vereins (random intercept, fixed slope)

$$y_{ij} = b_{0j} + b_1 \cdot x_{ij} + r_{ij} \; (i = 1, \dots, n_j; j = 1, \dots, m)$$



- die drei Vereine unterscheiden sich darin, wie hoch das Leistungsvermögen bei Stress = 0 (Intercept) ist, daher random intercept, weil der Wert unterschiedlich sein darf
- der Zusammenhang zwischen Prädiktor und Kriterium ist aber in allen Vereinen gleich (fixed slope b_1)

Quelle: Midway (o.D.), Figure 9.1 Link

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127ff)

Berücksichtigung unterschiedlicher Regressionsanstiege in Abhängigkeit des Vereins (random intercept, random slope)

$$\boxed{y_{ij} = b_{0j} + b_{1j} \cdot x_{ij} + r_{ij} \; (i = 1, \dots, n_j; j = 1, \dots, m)} \quad \text{mit} \quad \boxed{b_{1j} = b_1 + u_{1j}}$$

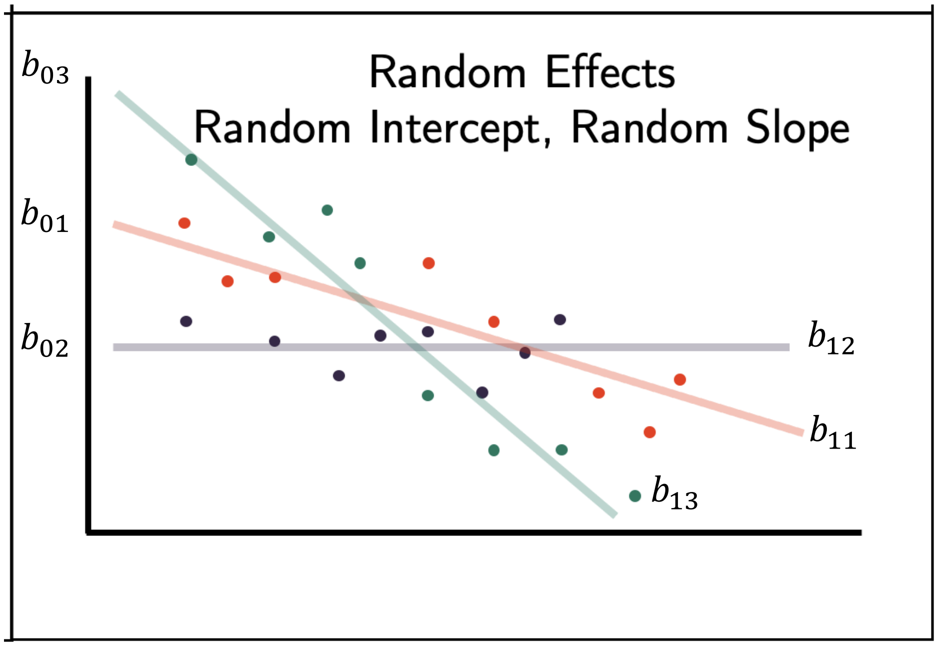
y_{ij}: Wert der Kriteriumsvariablen y des i-ten Spielers im j-ten Verein
x_{ij}: Wert der Prädiktorvariablen x des i-ten Spielers im j-ten Verein
r_{ij}: Residuum des i-ten Spielers im j-ten Verein
b_{0j}: Regressionskonstante (Intercept) für den j-ten Verein
u_{0j}: Abweichung der Regressionssteigung des j-ten Vereins von der mittleren Steigung
b_{1j}: Regressionssteigung (Slope) für den j-ten Verein
n_j: Anzahl der Spieler im j-ten Verein
m: Anzahl der Vereine

- Slope b_{1j}:
  - Unterschiedliche „Steigungswinkel“ für die unterschiedlichen Vereine sind möglich → Zusammenhang zwischen Prädiktor und Kriterium variabel

Feste und zufällige Effekte in Mehrebenenmodellen (z. B. Rudolf & Vogel-Blaschka, 2023, S. 127ff)

Berücksichtigung unterschiedlicher Regressionsanstiege in Abhängigkeit des Vereins (random intercept, random slope)

$$y_{ij} = b_{0j} + b_{1j} \cdot x_{ij} + r_{ij} \; (i = 1, \ldots, n_j; j = 1, \ldots, m)$$



- die drei Vereine unterscheiden sich darin, wie hoch das Leistungsvermögen bei Stress = 0 (Intercept) ist, daher random intercept, weil der Wert unterschiedlich sein darf
- der Zusammenhang zwischen Prädiktor und Kriterium (Slope b_{1j}) ist bei den drei Vereinen unterschiedlich

Quelle: Midway (o.D.), Figure 9.1 Link

Vorgehen bei der Mehrebenenmodellierung

- Ausführlichkeit und spezifische Arbeitsschritte unterscheiden sich von Fall zu Fall, aber folgende Struktur dient zur Orientierung

- Visuelle Inspektion der Daten (Scatterplots, Boxplots).
- Berechnung der Intraclass-Korrelationen (ICC) und des Design-Effekts zur Begründung der Mehrebenenmodellierung
- Deskriptive Beschreibung der Gruppestruktur (Gruppenmittel, Varianzen, Ausreißerwerte etc.)
- Berechnung von Mehrebenenmodellen und Finden des besten Modells für die Forschungsfrage (typischerweise Modellvergleiche)

Vorgehen bei der Mehrebenenmodellierung (z. B. Rudolf & Vogel-Blaschka, 2023, S. 132ff)

Berechnung der Intraclass-Korrelation (ICC)

- drückt den Anteil der Varianz in der Kriteriumsvariablen aus, der durch die Level-2-Variable erklärbar ist (bzw. allgemein durch Variablen höherer Level)
- Beispiel für zwei Ebenen → zwei Varianzquellen:
  - Level 1: Varianz innerhalb des Vereins (zwischen den Spieler/-innen)
  - Level 2: Varianz zwischen den Vereinen
- Berechnung des „Nullmodells“ („Null“, weil ohne Prädiktoren)
  - Level 1: $y_{ij} = b_{0j} + r_{ij}$ $(i = 1, \dots, n_j; j = 1, \dots, m)$
  - Level 2: $b_{0j} = g_{00} + u_{0j}$ $(j = 1, \dots, m)$
- Einsetzen der Level-Gleichung in Level 1-Gleichung:

$$\boxed{y_{ij} = g_{00} + u_{0j} + r_{ij} \ (i = 1, \dots, n_j; j = 1, \dots, m)}$$

Vorgehen bei der Mehrebenenmodellierung (z. B. Rudolf & Vogel-Blaschka, 2023, S. 132ff)

Berechnung der Intraclass-Korrelation (ICC)

$$\boxed{y_{ij} = g_{00} + u_{0j} + r_{ij} \; (i = 1, \ldots, n_j; j = 1, \ldots, m)}$$

- Gleichung entspricht einer „Varianzzerlegung“ bezüglich unserer Varianzquellen:
  - g_{00}: Gesamtmittelwert der gesamten Stichprobe (erwarteter Wert ohne Prädiktoren) → Varianz = 0, da konstanter Wert
  - u_{0j}: gruppenspezifische Abweichung vom Gesamtmittelwert in der Gruppe j → Varianz auf Level 2: $\sigma^2_{u_0}$
  - r_{ij}: individuelle Abweichung des Falls i in Gruppe j vom gruppenspezifischen Erwartungswert → Varianz auf Level 1: σ^2_r

- Berechnung der ICC: $$\boxed{ICC = \frac{\sigma^2_{Level\ 2}}{\sigma^2_{Level\ 1} + \sigma^2_{Level\ 2}} = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_{u_0}}}$$

Vorgehen bei der Mehrebenenmodellierung

Berechnung der Intraclass-Korrelation (ICC)

$$\boxed{ICC = \frac{\sigma^2_{Level\ 2}}{\sigma^2_{Level\ 1} + \sigma^2_{Level\ 2}} = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_{u_0}}}$$

- Interpretation des ICC-Werts: je größer, desto stärker ist der Anteil der Varianz durch Gruppenunterschiede auf Level 2 entsteht
  - Orientierung: ab einem ICC von .05 (5 %) sollte man und ab .10 (10 %) muss man definitiv ein Mehrebenenmodell nutzen; **aber** selbst bei kleineren Werten (z. B. .01; 1%) können die Standardfehler in normalen Regressionen schon verzerrt sein, wenn die Gruppen sehr groß sind
  - Unabhängig von der Größe: Selbst wenn der ICC klein ist, rechnet man trotzdem ein MLM, wenn das Forschungsinteresse eine Cross-Level Interaction (also einen Moderationseffekt von Level 2 auf Level 1) ist
    → man kann so sicherstellen, dass man die Struktur der Daten korrekt berücksichtigt hat

Vorgehen bei der Mehrebenenmodellierung

Berechnung des Design-Effekts

- Design-Effekt drückt aus, um wie viel sich die Fehlerschätzung vergrößert, weil die Daten nicht unabhängig voneinander sind, sondern in Gruppen „verklumpt“ sind
  - Wichtig: bezieht sich auf Abhängigkeit durch Struktur, nicht Unterschiede durch Eigenschaften (z. B. Geschlecht)

- Berechnung: $DEFF = 1 + (\bar{n} - 1) \cdot ICC$
- Beispiel: 20 Vereine mit durchschnittlich 10 Spielern und ICC ist 10%
- $DEFF = 1 + (10 - 1) \cdot 0.10 = 1{,}9$ → Standardfehler ist $\sqrt{1{,}9} \approx 1{,}38$ mal größer als bei einer normalen Regression oder anders gesagt:
  Die 200 Spieler liefern nur so viel statistische Power wie (200/1.9) ungefähr 105 einzelne, unabhängige Spieler.

Vorgehen bei der Mehrebenenmodellierung

Berechnung der Modelle (idealtypisches Vorgehen)

- Nullmodell
  - (siehe oben → Gibt es überhaupt etwas auf Gruppenebene?)
- Random-Intercept-Modell (Level-1-Prädiktoren)
  - isoliert within-Effekte und kontrolliert die Cluster-Abhängigkeit → Baseline-Erklärmodell für Unterschiede innerhalb der Gruppen
- Means-as-Outcome-Modell (Level-2-Prädiktoren auf Intercept)
  - erklärt, warum sich die Gruppen im Ausgangsniveau unterscheiden
- Random-Coefficients-Modell (Random Slopes)
  - prüft, ob Einfluss von Prädiktor auf Kriterium in allen Gruppen gleich ist (**Achtung**: Modellierung muss theoretisch begründbar sein!)
- Cross-Level-Interaction-Modell (Einfluss von Level 2 auf Level 1)
  - erklärt durch Einfluss von Level 2 die unterschiedliche Wirkung von Prädiktor auf Kriterium in den verschiedenen Gruppen

Vorgehen bei der Mehrebenenmodellierung

Berechnung der Modelle (idealtypisches Vorgehen)

- Anmerkungen:
  - Man „darf“ auch gleich komplexere Modelle erstellen statt all diese Schritte nacheinander zu durchlaufen.
  - Es dürfen auch Kovariaten auf allen Ebenen mitmodelliert werden, allerdings sollten diese immer theoretisch begründet sein, da diese die Modellierung komplexer machen und Freiheitsgrade „verbrauchen“.
  - Es gilt wie bei allen Regressionen das Prinzip der Sparsamkeit: Mehr Parameter im Modell müssen eine statistisch bedeutsame Verbesserung der Modellgüte (siehe nächste Folie) mit sich bringen.
  - Im Rahmen der Modellberechnungen erfolgt typischerweise eine Zentrierung der Werte (siehe unten), um within- und between-Effekte sauber trennen zu können.
  - Modellvergleich dient nicht dazu, „so lange zu probieren, bis wir das beste haben“, sondern folgt inhaltlich-sachlogischen Prinzipien → generell schlägt dabei theoretische Begründbarkeit statistische Ergebnisse

Modellgüte bei Mehrebenenmodellierung (Field, 2018)

- Test des Overall-Fit mittels Chi-Quadrat-Test des -2LL (siehe Sitzung „Logistische Regression“
- Weitere mögliche Kennzahlen: AIC, AICC, CAIC, BIC → keine generelle Interpretation; nur Vergleich zwischen den einzelnen Modellen wenn jeweils ein neuer Parameter hinzukommt
- **Vorsicht**: Als Schätzmethode muss Maximum-Likelihood verwendet werden, wenn man Modelle auf Basis dieser Werte vergleichen möchte
  - Restricted Maximum Likelihood (REML) wird für das finale Modell zur Schätzung benutzt
- Vorgehen folgt dieser grundsätzlichen Logik:
  - Basismodell → Fit
  - Erweiterung um eine Variable → Fit → signifikante Verbesserung? → wenn ja, dann nächstes Modell, wenn nein, dann Stop

Voraussetzungen bei Mehrebenenmodellierung (Field, 2018)

- bezüglich der random effects: Residuen auf Gruppenlevel sind normalverteilt bezogen auf das Gesamtmodell
  - Beispiel: unterschiedliche Intercepts der Gruppen streuen normalverteilt um den durchschnittlichen Intercept

- Multikollinearität erzeugt Probleme → Zentrierung der Prädiktoren schafft Abhilfe
  - Group Mean Centring, Grand Mean Centring oder Kombination aus beiden je nach Kontext und Daten

- die strengen (in der Praxis kaum vollständig erfüllbaren) Voraussetzungen für die Repeated-Measures-ANOVA entfallen, daher wird ab 3 Messzeitpunkten empfohlen, MLM zu nutzen (Level 2: Proband/in, Level 1: Messungen)

Zentrierung bei Mehrebenenmodellierung (Field, 2018)

zwei Varianten, Anwendung je nach Forschungsinteresse

- Grand-Mean Centering
  - von jedem Wert wird der Gesamtdurchschnitt aller Teilnehmer abgezogen
  - Anwendung, wenn die Multikollinearität gesenkt werden soll oder die Variable als reine Kontrollvariable (Kovariate) dient.
  - Effekt: Intercept des Modells ist nun der Wert für eine "durchschnittliche Person" aus der gesamten Stichprobe → erleichtert die Interpretation der Ergebnisse, weil "Null" jetzt "Durchschnitt" bedeutet.
- Group-Mean Centering
  - von jedem Wert den Mittelwert seiner jeweiligen Gruppe abgezogen
  - Anwendung, wenn der reine individuelle Effekt interessiert, bereinigt um alles, was die Gruppenzugehörigkeit betrifft
  - Effekt: Unterschiede zwischen den Gruppen aus dieser Variable werden „herausgelöscht“ → verhindert, dass sich der Level-1-Effekt mit dem Level-2-Effekt vermischt

Zentrierung bei Mehrebenenmodellierung (Field, 2018)

zwei Varianten, Anwendung je nach Forschungsinteresse

- Take Home Message
  - Grand Mean Centring: Wie schneidet eine VP im Vergleich zum Durchschnitt aller VP ab?
    - Beispiel: Spieler hat eine Motivation von 7, durchschnittliche Motivation über alle Spieler ist 5 → +2 (Motivation ist überdurchschnittlich)
  - Group Mean Centring: Wie schneidet eine VP im Vergleich zu den anderen VP in seiner Gruppe ab?
    - Beispiel: Spieler hat eine Motivation von 7, durchschnittliche Motivation in seinem Verein ist 8 → -1 (Motivation ist unterdurchschnittlich)
  - Wenn Cross-Level Interactionen (Level 2 wirkt auf Level 1) betrachtet werden, ist Group-Mean Centering oft die sauberere Wahl, weil es den Interaktionseffekt "isoliert".
  - Level-2-Variablen können nur am Grand Mean zentriert werden, da es keinen Gruppenmittelwert gibt, der darüber liegt (außer, es gibt Level-3-Variablen).

Beispiele für Moderation bzw. Mediation Fachzeitschriften

their reported levels of anxiety. In two representative samples of Finnish school students ($N = 678$ schools/71,392 students; $N = 704$ schools/85,989 students), weak identification environment was related to increased anxiety. In addition, in schools where identification environment was weaker, the student level relationship between perceived physical environment and anxiety was stronger, and students were more anxious. Our results provide evidence that identification

Beispiel für die Ergebnisse von ICC und DEFF einer Studie mit Mehrebenenmodell
Quelle: Finell et al. (2024)

First, we estimated a null model for each variable. In a null model, there is only one variable – variance at student and school levels and the ICC. The ICC reports the proportion of variance that belonged to the school level (Hox, 2010). Then we calculated the design effect (DEFF) of each variable (Table 2). This measure is used to estimate whether multilevel modelling is needed, that is if the DEFF of the outcome variable is greater than 1.1 (Lai & Kwok, 2015). DEFF is estimated as a function of the ICC and average cluster size (Lai & Kwok, 2015; Muthén & Satorra, 1995).[4]

Study results show that parental monitoring was an important predictor of academic achievement ($\gamma_{40} = 0.51, p < .001$), antisocial school behaviors ($\gamma_{40} = -28.12, p < .001$), and prosocial school behaviors ($\gamma_{40} = 23.50, p < .001$) across all schools after controlling for student demographics of gender, ethnic background, and SES background. Consistent with patterns found in prior research (e.g., Crouter et al., 1990; Kristjansson & Sigfusdottir, 2009; Shumow & Lomax, 2002), taking into account the signs of the mean effects parental monitoring (γ_{40}s) on schools outcomes (school behaviors and achievement), the level of parental monitoring significantly predicted growth (increase) in students' academic achievement and prosocial school behaviors, and decline (decrease) in student antisocial school behavior.

Beispiel für Cross-Level-Interactions | Quelle: Top, Liew, & Luo (2017)

Zusammenfassung I

Mehrebenenmodelle können …

- … Abhängigkeiten durch Ebenenstrukturen korrekt berücksichtigen (z. B. Personen in Gruppen, Messzeitpunkte in Personen)
- …Varianz auf verschiedenen Ebenen trennen (individuell vs. gruppenspezifisch)
- … Effekte innerhalb und zwischen Gruppen unterscheiden
- … erklären,
  - warum Gruppen unterschiedlich starten (Random Intercepts)
  - warum Effekte unterschiedlich wirken (Random Slopes)
- … Kontextabhängigkeiten modellieren (Cross-Level-Interaktionen)

Zusammenfassung II

Mehrebenenmodelle können **nicht** …

- fehlende Theorie ersetzen
- kausale Effekte garantieren
- schlechte Messungen „reparieren“
- beliebige Gruppenvariablen automatisch zu Clustern machen (z. B. Geschlecht ≠ Cluster)
- Modellkomplexität rechtfertigen, nur weil der Fit besser wird

Zusammenfassung III

- Daten sind oft nicht unabhängig
  → Mehrebenenmodelle berücksichtigen diese Abhängigkeit
- Das Nullmodell zeigt,
  ob Varianz auf Gruppenebene vorhanden ist (ICC, Design-Effekt)
- Die Modellierung erfolgt schrittweise:
  - Random Intercepts (Unterschiede zwischen Gruppen)
  - Level-1-Prädiktoren (Within-Effekte)
  - Level-2-Prädiktoren (Between-Effekte)
  - Random Slopes (unterschiedliche Effekte)
  - Cross-Level-Interaktionen (Kontexteffekte)
- Modellvergleiche dienen dem Verstehen, nicht dem „Best-Fit-Wettbewerb"
- Zentrale Frage ist immer: Welche Annahmen über die Datenstruktur sind theoretisch plausibel?

Prüfungsliteratur

- Rudolf, M., & Vogel-Blaschka, D. (2023). *Komplexe Regressionsanalytische Verfahren. Eine praxisorientierte Einführung mit Anwendungsbeispielen in R und SPSS*. Hogrefe.
  - Mehrebenenmodelle (S. 121–222)

Weiterführende Literatur

- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th Edition). Sage.
  - Multilevel Linear Models (S. 935–990)
- Hox, J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel Analysis. Techniques and Applications* (3rd Edition). Routledge.
- Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society. Series B (Methodological), 13*(2), 238–241. http://www.jstor.org/stable/2984065