10. Jahrestagung des Arbeitskreises Evaluation und Qualitätssicherung Berliner und Brandenburger Hochschulen
Lehre und Studium professionell evaluieren: Wie viel Wissenschaft braucht die Evaluation?
Universität Potsdam, 26.03.-27.03.2009

Forum 5: Evaluation und fortgeschrittene Analyseinstrumente

MEHREBENENANALYSE: ANGEMESSENE MODELLIERUNG VON EVALUATIONSDATEN

Potsdam 27. März 2009

Dr. Manuela Pötschke

HINTERGRUND

Neue Anforderungen an Evaluation (hier aus Fachbereichssicht)

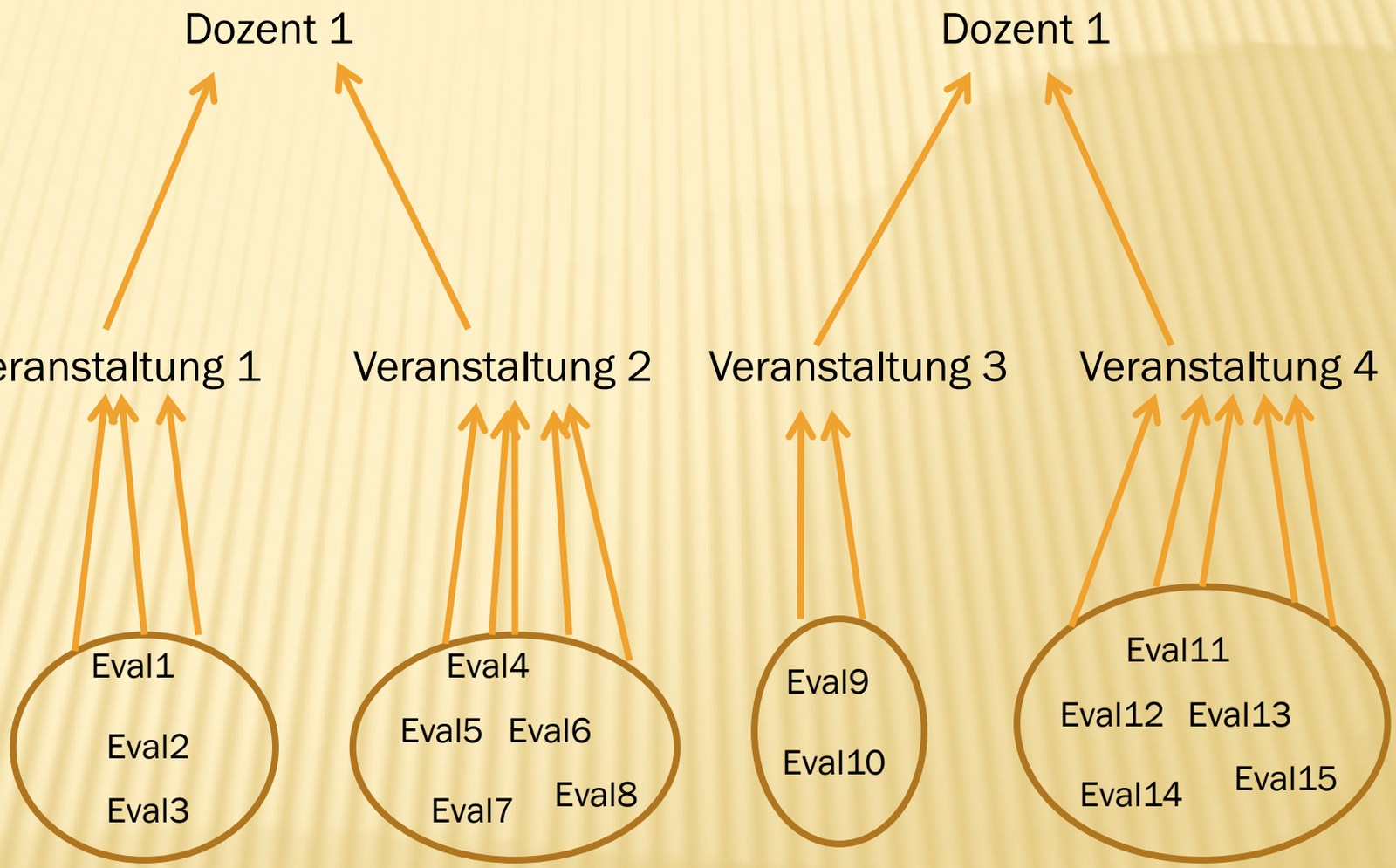
- Berichterstattung
- Mittelzuweisung
- Flächendeckende Evaluationen
- Konsequenzen?
- Bezug der Beurteilung?
- Kriterien der Beurteilung?
- Komplexität der erhobenen Daten

HIERARCHISCHE DATENSTRUKTUR

3. Ebene:
Dozent (N=60)

2. Ebene:
Veranstaltung (N=150)

1. Ebene:
studentisches
Urteil (N≈2300)



DAS EMPIRISCHE BEISPIEL: ERKLÄRUNG INDIVIDUELLER LEHREVALUATIONSURTEILE

Unabhängige Variablen

Dozent (N=60)
- Geschlecht
- Statusgruppe: Professor, Promovierter Mitarbeiter, nicht promovierter Mitarbeiter
- Persönlichkeitseigenschaften: Sympathie, Kreativität, Offenheit, Attraktivität, Kompetenz (N=30)

Veranstaltung (N=150)
- Fach: Soziologie, Politikwissenschaft, Geographie, Religionswissenschaft, Master Politik
- Veranstaltungstyp: Vorlesung, Seminar, Übung

studentisches Urteil (N≈2300)
- Aussagen zur Veranstaltung, dem Stoff und dem Dozent
- Aussagen zur Lernatmosphäre und dem Verhältnis zu anderen
- Eigene Leistungsbeurteilung
- Geschlecht, Fach

Abhängige Variable

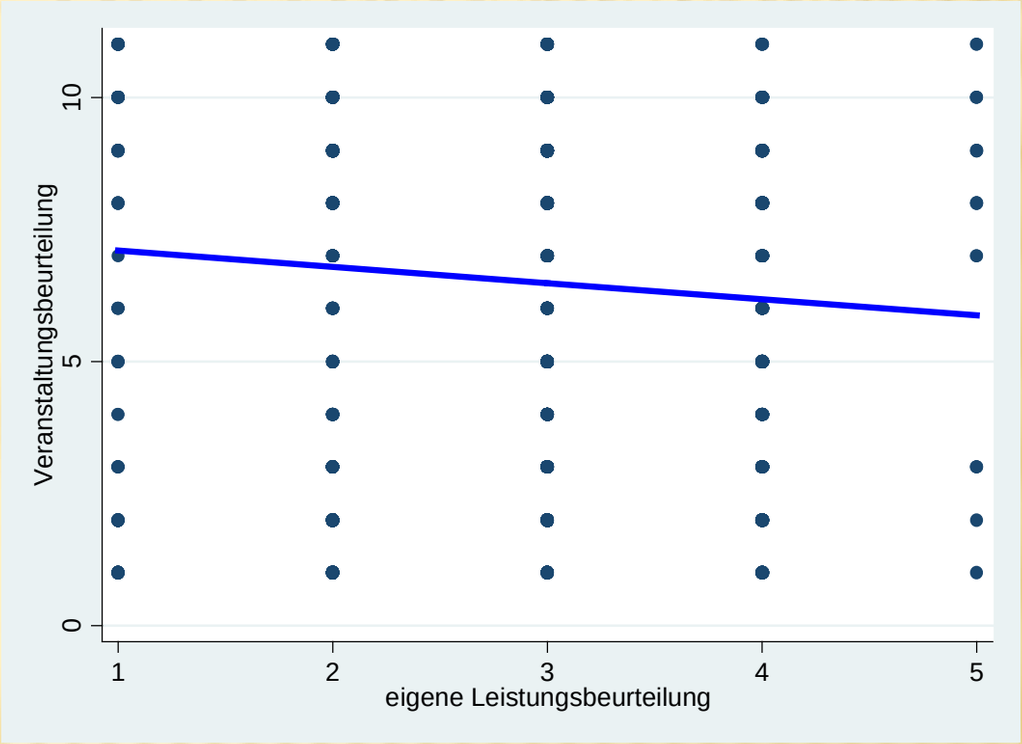
Studentisches Evaluationsurteil

VERNACHLÄSSIGUNG DER DATENSTRUKTUR

Beispiel: Wie wirkt sich die eigenen Leistungsbeurteilung auf die Veranstaltungsbeurteilung aus?

Veranstaltungsbeurteilung: Alles in allem, wie beurteilen Sie die Qualität der Veranstaltung insgesamt? (1=sehr schlecht bis 11=sehr gut)

Leistung: eigene Leistungsbeurteilung der Studierenden (1=sehr gut bis 5= sehr schlecht)



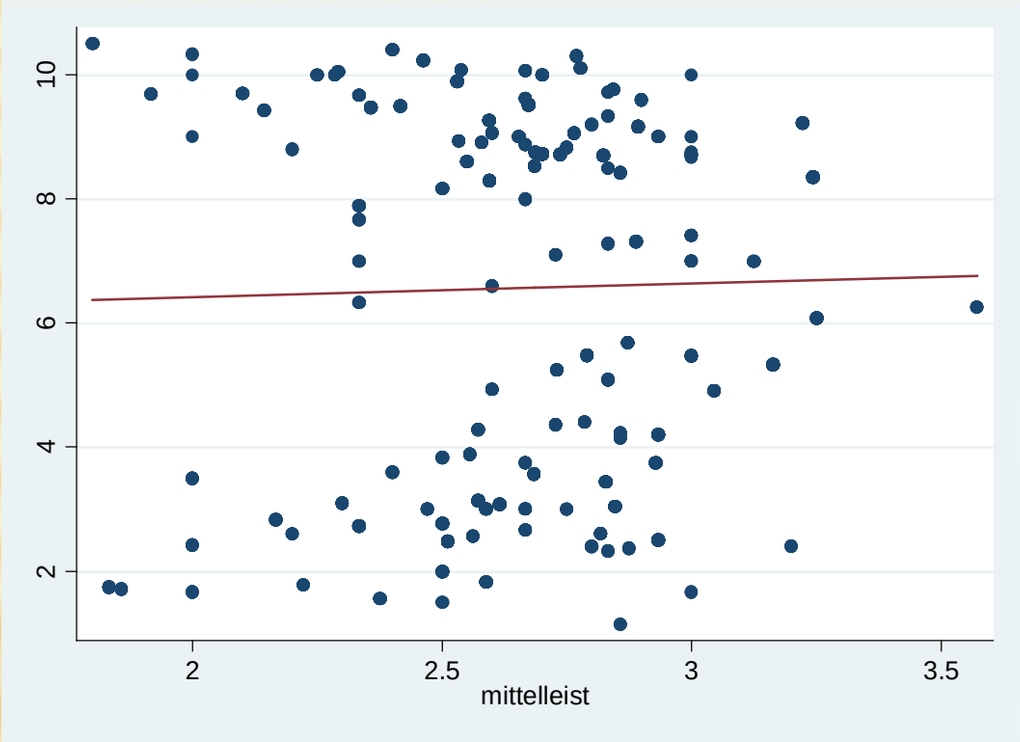
$R^2=0{,}0048$ $y=7{,}411-0{,}3082x+e$ (***)

Probleme:

Unterstellung der Unabhängigkeit aller Beurteilungen der Studierenden über die Veranstaltungen und Dozenten hinweg

globales und wenig aussagekräftiges Bild, durch die ausschließliche Orientierung am Gesamtmittelwert über alle Veranstaltungen und alle Dozenten hinweg

AGGREGIERUNG DER INDIVIDUALDATEN



R^2=0,0005 (n.s.)

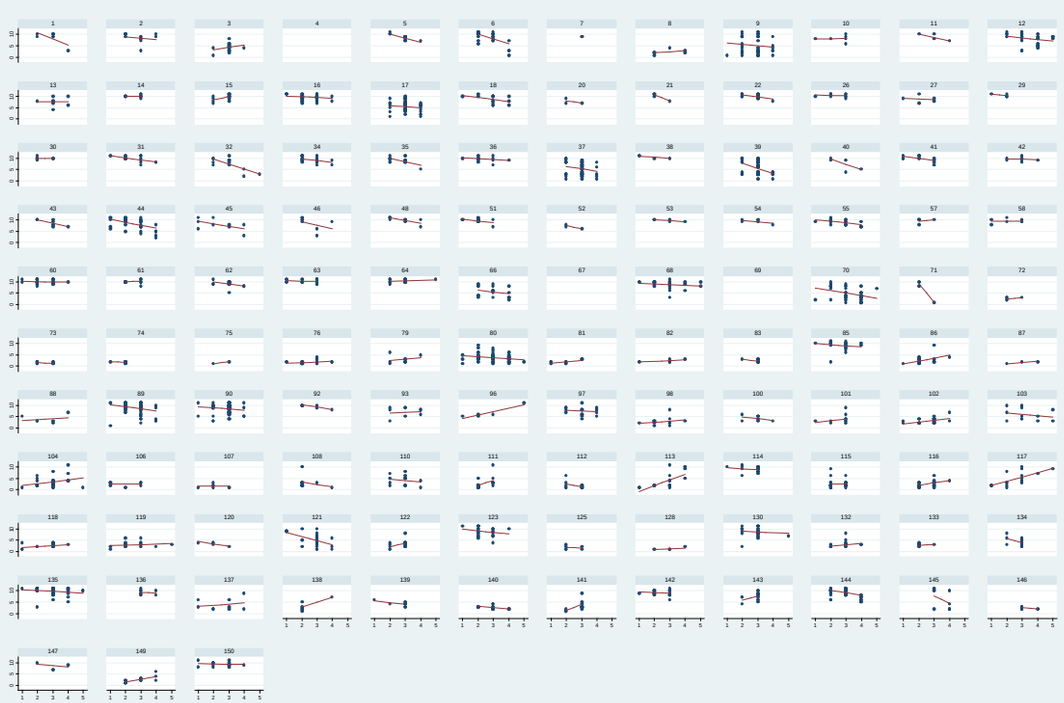
y=5,963+ 0,2235x+ e

Probleme:

- Durch die Aggregierung der Daten geht viel Information, insbesondere aber die Varianzen innerhalb der Aggregateinheiten verloren.
- Die Aggregate werden im Rahmen eines solchen Vorgehens als in sich homogen betrachtet, was extrem fragwürdig ist.
- Die Ergebnisse auf der Aggregatebene weichen von den Ergebnissen auf der Individualebene teilweise erheblich ab (sog. aggregation bias)
- Die Untersuchung des Zusammenwirkens von Individual- und Aggregatvariablen ist nicht möglich

- Die Ergebnisse einer Aggregatanalyse lassen sich theoretisch nur schwer interpretieren. Jede Kausalinterpretation muss eigentlich auf individueller Ebene argumentieren, die Information über diese wurde durch die Aggregation jedoch fallen gelassen.

DURCHFÜHRUNG SEPARATER REGRESSIONSANALYSEN



Probleme:

- Hier wird zwar zugelassen, dass in den verschiedenen Aggregateinheiten unterschiedliche Beziehungsstrukturen existieren, doch ist ein solches Vorgehen nur bei einer kleinen Zahl von Level-2-Einheiten praktikabel.
- Es ist im Rahmen dieses Vorgehens nicht möglich, den Einfluss von Aggregatvariablen auf die abhängige Variable (Kontexteffekt) und die Stärke des Effekts von Mikrovariablen (Cross-Level-Interaktion) zu untersuchen.

„SLOPES AS OUTCOMES“-ANALYSEN

Modellidee:

Nachdem für jede Einheit der 2. Ebene ein separates Regressionsmodell geschätzt wurde, werde die Regressionsparameter Intercept und Slope aus diesen Regressionsmodellen durch Merkmale der 2. Ebene erklärt.

Probleme:

- Die Koeffizienten der Regressionsmodelle innerhalb der Aggregateinheiten basieren in der Regel auf sehr geringen Fallzahlen. Die Schätzwerte für Regressionskonstante und -koeffizienten sind daher nicht sehr reliabel. Wenige „Ausreißer“ können eine erhebliche Fehlervarianz bewirken („bouncing betas“)
- Angemessene Mehrebenenmodelle müssen daher die Stichprobengrößen, auf denen die Schätzer der Koeffizienten basieren, berücksichtigen.
- Keine Analyse von Cross-Level-Interaktionen möglich
- Die „Slopes as Outcomes“-Analyse ist in der Regel beschränkt auf die Analyse *einer* unabhängigen Mikro-Variablen.

BEGRIFF UND IDEE DER MEHREBENENANALYSE

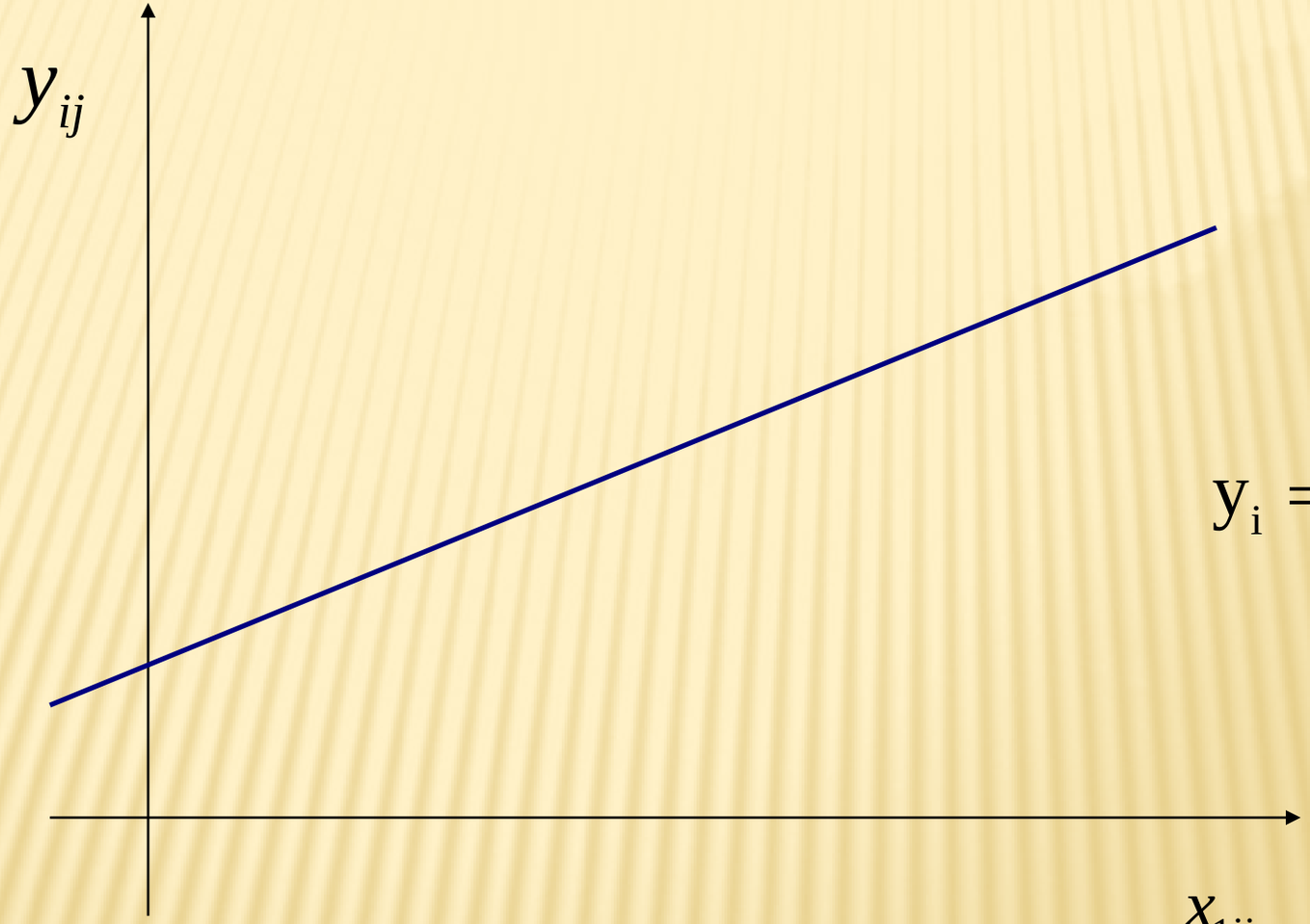
Mehrebenenanalysen als angemessenes Verfahren:

- spiegelt die Datenstruktur wieder und kommt deshalb zu korrekten Schätzungen
- erlaubt die Analyse und Darstellung komplexer Datenstrukturen

Inhaltliche Begründung aus empirischen Studien und theoretischen Überlegungen
Methodische Begründungen

Man spricht immer dann von einer Mehrebenenanalyse, wenn „Objekte verschiedener Ordnung gleichzeitig zum Gegenstand der Analyse werden“ (Hummell 1972: 13) und ihre Wirkung auf eine abhängige Variable simultan bestimmt wird.

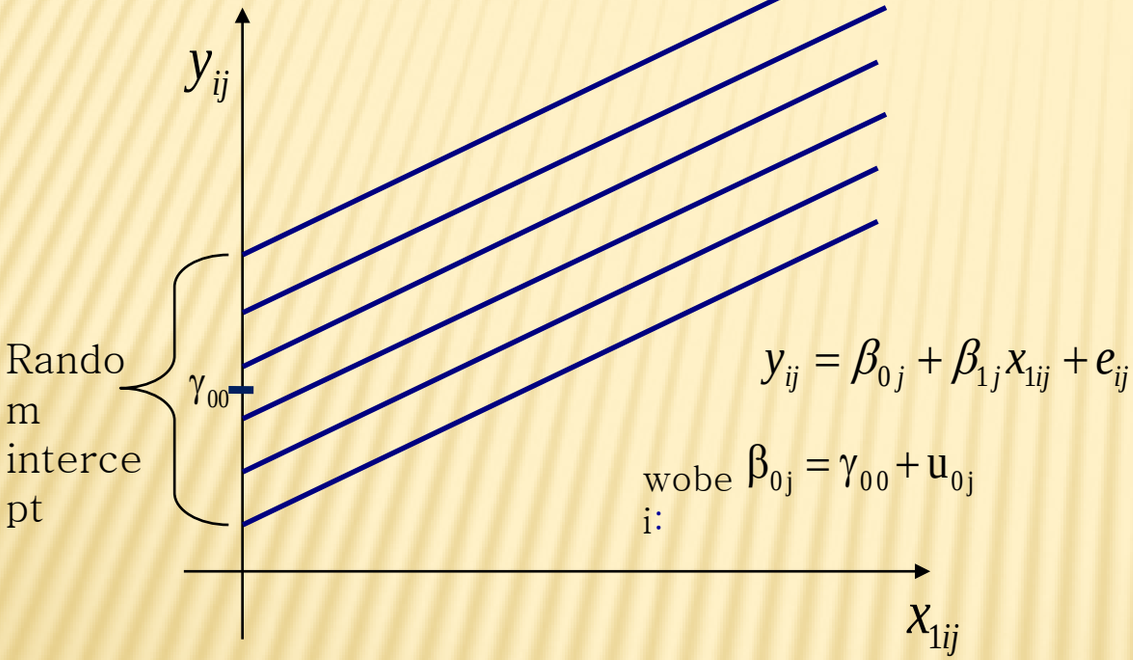
LINEARE EINFACHREGRESSION



$$y_i = \beta_0 + \beta_1 x_{1i} + e_i$$

$$\tau_0 = 0 \qquad \tau_1 = 0$$

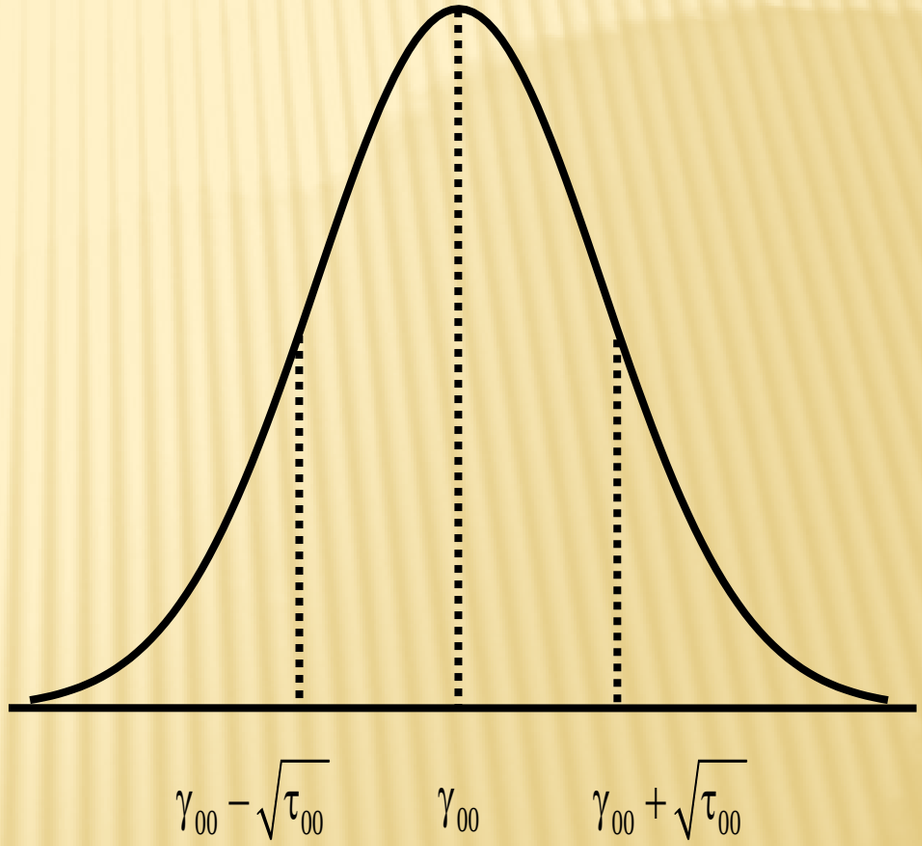
RANDOM INTERCEPT



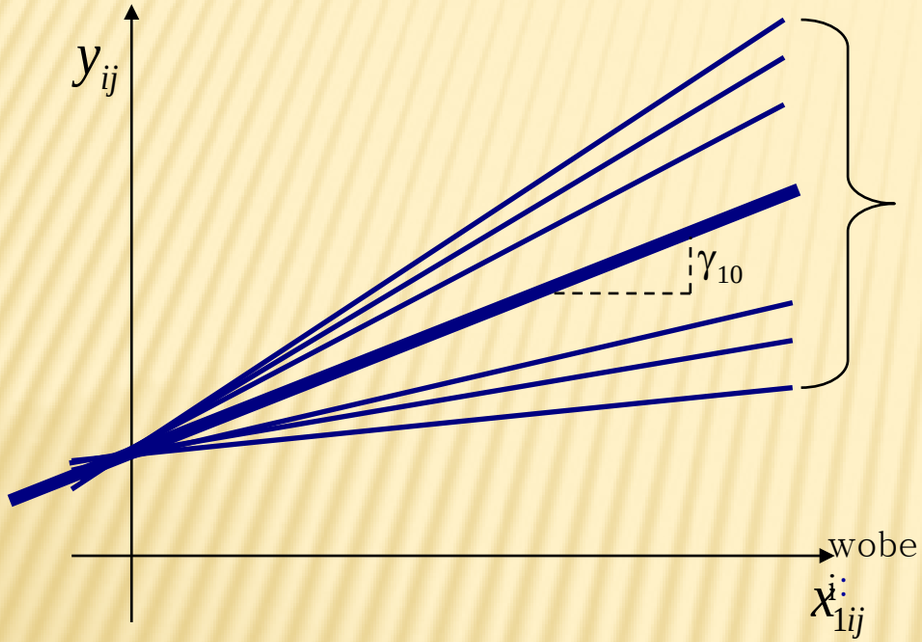
$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

wobei: $\beta_{0j} = \gamma_{00} + u_{0j}$

$$\tau_0 \neq 0 \qquad \tau_1 = 0$$



RANDOM SLOPES

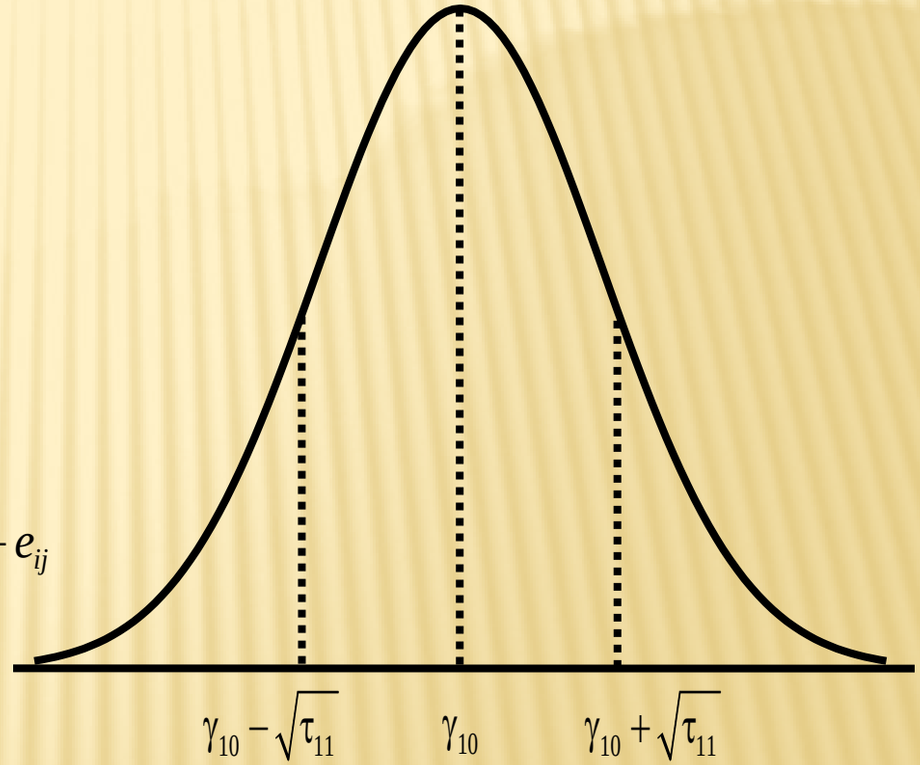


Random slopes

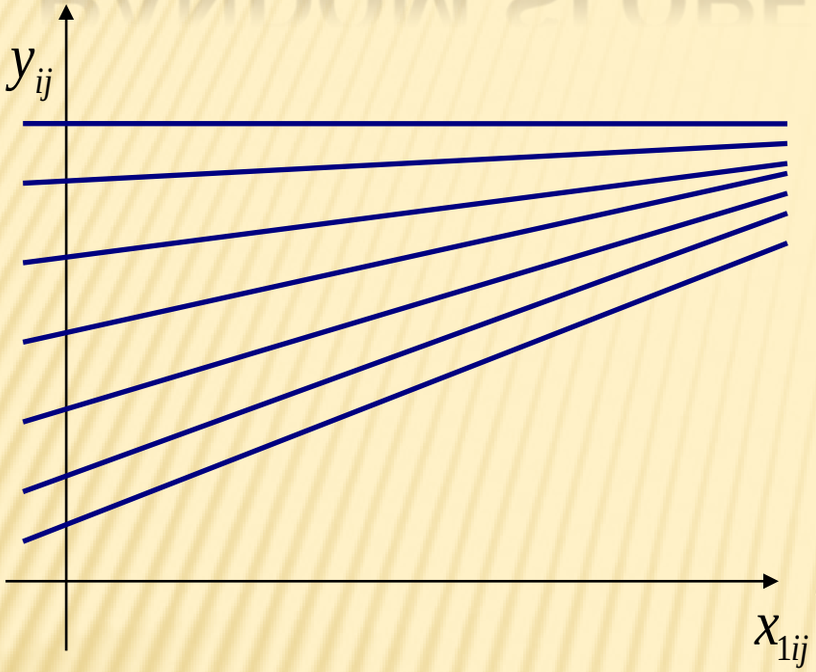
$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\tau_0 = 0 \quad \tau_1 \neq 0$$



RANDOM SLOPE AND RANDOM INTERCEPT



$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

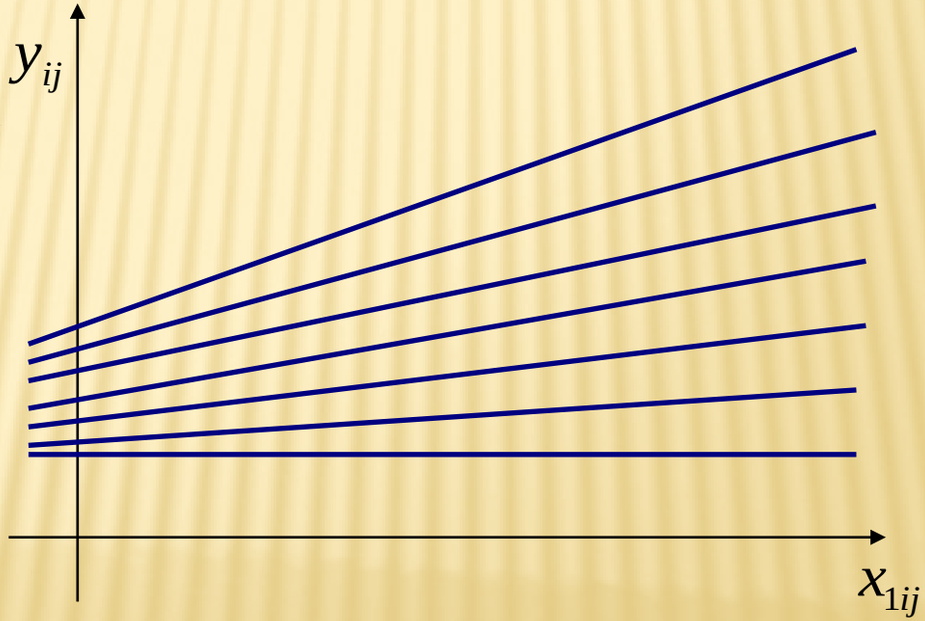
wobei

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

und

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$\tau_0 \neq 0$ $\tau_1 \neq 0$ $\tau_{01} < 0$



$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

wobei

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

und

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$\tau_0 \neq 0$ $\tau_1 \neq 0$ $\tau_{01} > 0$

MIXED FORM

Einfügen von …

$$\beta_{0j} = \beta_0 + u_{0j} \quad \text{und} \quad \beta_{1j} = \beta_1 + u_{1j}$$

in …

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

$$i = 1,\ldots, n$$

$$j = 1,\ldots, J$$

ergibt…

$$y_{ij} = \beta_0 + u_{0j} + \beta_1 x_{1ij} + u_{1j} x_{1ij} + e_{ij}$$

umgestellt:

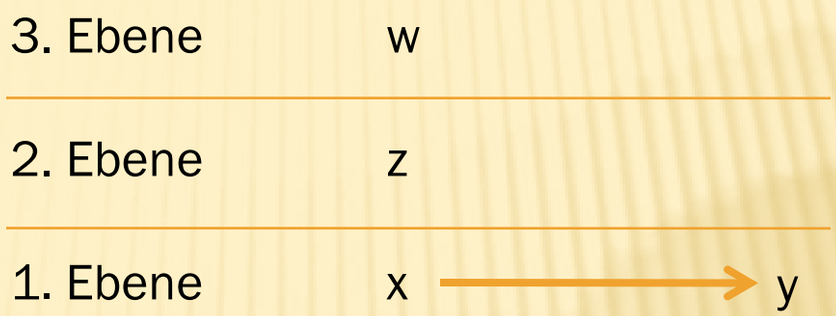
$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij}$$

‚Fixed Part'

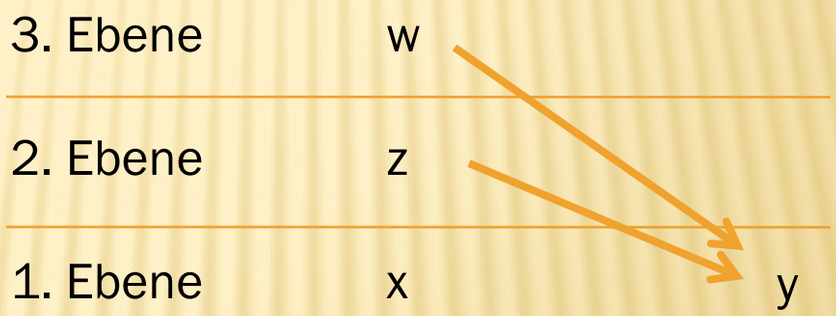
‚Random Part'

MÖGLICHE EFFEKTE

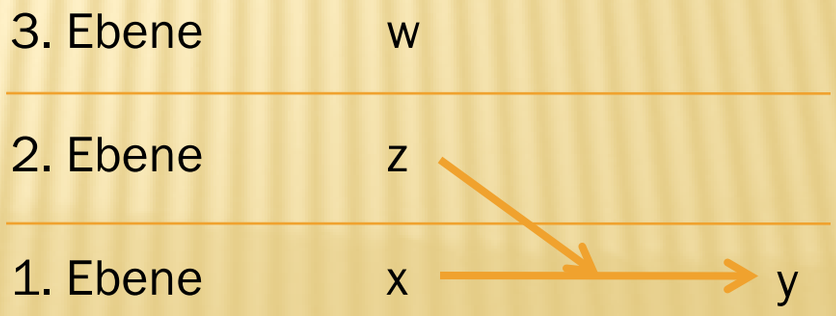
Effekte von Individualvariablen



Effekte von Aggregatvariablen



Cross-Level-Interaktionen



WARUM MEHREBENENANALYSEN? POTENTIALE

- die Zerlegung der Gesamtvarianz in die durch die einzelnen Ebenen erklärbaren Varianzanteile.

Empty Model

- die gleichzeitige Untersuchung von Effekten unterschiedlicher Merkmale aus verschiedenen Aggregatebenen.

Random Intercept Model

- den Einbezug von Wechselwirkungen über unterschiedliche Aggregatebenen hinweg (Cross-Level-Wechselwirkungen).

Cross Level Effect Model

- die gleichzeitige Berücksichtigung zufälliger Variation auf unterschiedlichen Ebenen.

Random Coefficient Model

- die korrekte Schätzung von Standardfehlern in mehrstufigen Zufallsauswahlen.

Designeffekt

NULLMODELL: EMPIRISCHES BEISPIEL

Leeres Modell ohne erklärende Variable(n)

$$\text{Veranstaltungsevaluation}_{ijk} = \beta_0 + r_{0k} + u_{0jk} + e_{ijk}$$

Varianzzerlegung auf die Ebenen:

Dozent: 10,22 %
Veranstaltung: 60,64 %
Studierende: 29,14 %

var(r) = 1,2609
τ_{00} = 7,4835
σ^2_e = 3,5968

Mittlere Schätzung über alle Personen in allen Gruppen hinweg: 6,42

Modellgüte: -2ll=8822,3916

VORLÄUFIGE EMPIRISCHE ERGEBNISSE

Erklärungen durch Variablen der ersten Ebene

Über die Ziele wurde gut informiert. (fünfstufig: stimme nicht zu - stimme zu)	0,293***	var(r) = 0,8952 τ_{00} = 8,7417 σ^2_e = 3,242
Die Basisliteratur verstehe ich gut. (fünfstufig: stimme nicht zu - stimme zu)	0,161***	
Die Studierenden wurden gut betreut. (fünfstufig: stimme nicht zu - stimme zu)	0,364***	
Atmosphäre insgesamt (fünfstufig: sehr gut – sehr schlecht)	-0,402***	
Ich bin schnell abgelenkt (fünfstufig: stimme nicht zu - stimme zu)	-0,250***	-2ll = 6712,437*** (zum Nullmodell)

VORLÄUFIGE EMPIRISCHE ERGEBNISSE

Erklärungen durch Variablen der zweiten Ebene

Über die Ziele wurde gut informiert. (fünfstufig: stimme nicht zu - stimme zu)	0,293***
Die Basisliteratur verstehe ich gut. (fünfstufig: stimme nicht zu - stimme zu)	0,163***
Die Studierenden wurden gut betreut. (fünfstufig: stimme nicht zu - stimme zu)	0,366***
Atmosphäre insgesamt (fünfstufig: sehr gut – sehr schlecht)	-0,406***
Ich bin schnell abgelenkt (fünfstufig: stimme nicht zu - stimme zu)	-0,250***
Referenz Master Politik	
Soziologie	5,1***
Politikwissenschaft	3,898***
Geographie	4,923*
Religion	3,533**

var(r) = 1,5149

τ_{00} = 7,6337

σ^2_e = 3,2415

Kein signifikanter Einfluss des Veranstaltungstyps, wenn um die Atmosphäre kontrolliert wird.

-2ll = 6691,9616 ***
(zum letzten Modell und zum Nullmodell)

VORLÄUFIGE EMPIRISCHE ERGEBNISSE

Erklärungen durch Variablen der dritten Ebene

Über die Ziele wurde gut informiert. (fünfstufig: stimme nicht zu - stimme zu)	0,293***
Die Basisliteratur verstehe ich gut. (fünfstufig: stimme nicht zu - stimme zu)	0,163***
Die Studierenden wurden gut betreut. (fünfstufig: stimme nicht zu - stimme zu)	0,367***
Atmosphäre insgesamt (fünfstufig: sehr gut – sehr schlecht)	-0,406***
Ich bin schnell abgelenkt (fünfstufig: stimme nicht zu - stimme zu)	-0,250***
Referenz Master Politik	
Soziologie	5,342***
Politikwissenschaft	3,732***
Geographie	5,732**
Religion	3,58**
Referenz Assistent	
Professor	1,474*
Nicht promovierter Mitarbeiter	0,627

var(r) = 1,2269

τ_{00} = 7,7392

σ^2_e = 3,2412

Keine signifikanten Einflüsse des Geschlechts des Dozenten.

Keine signifikanten Einflüsse der Persönlichkeitsmerkmale der Dozenten.

-2ll = 6686,1536 ***
(zum Nullmodell)

FAZIT

Potentiale

- Mehrebenenmodelle sind der Evaluationsdatenstruktur angemessen
- Inhaltlich relevante zusätzliche Informationen für größere Einheiten
- Informationen sind übersichtlich zu präsentieren
- Einflüsse der Struktur und individueller Merkmale werden gleichzeitig einbezogen

Probleme mit/ in Mehrebenenmodellen

- Hohe Anforderungen an die Struktur der Daten und die Stichprobenziehung
- Konsequenzen der Verletzung von Modellannahmen noch nicht ausreichend untersucht
- Inhaltliche gehaltvolle Variablen auf unterschiedlichen Aggregatebenen
- Komplexität der impliziten Modellbeziehungen

Inhaltliche Schlussfolgerungen

- Hinweise auf angemessene Schätzungen durch Mehrebenenmodelle, die erweiterte Interpretationen zulassen
- Ambivalente Befunde zu den Einflüssen von Persönlichkeitseigenschaften der Dozenten bleiben bestehen
- Keine Unterschiede in der Beurteilung von Frauen und Männern