



Institut zur Qualitätsentwicklung
im Bildungswesen



Längsschnittliche Analysen

Sebastian Weirich und Nicklas Hafiz

Institut zur Qualitätsentwicklung im Bildungswesen (IQB)
Humboldt-Universität zu Berlin

Gesis Workshop, Oktober 2024



Längsschnittliche Analysen als GLMM



- Modellierung in der 1PL-Welt

- Zeit kann als Within-Person-Faktor (Längsschnitt) oder als Between-Person-Faktor (Kohortenquerschnitt) modelliert werden
- Aber: zu den bisher betrachteten Voraussetzungen (Raschhomogenität, lokale stochastische Unabhängigkeit, Invarianz des Messmodells [DIF]) kommt eine weitere hinzu
 - DIF: differentiell Itemfunktionieren (zwischen Personengruppen, bspw. Jungen vs. Mädchen)
 - differentiell Itemfunktionieren (zwischen Messzeitpunkten) wird auch als “Item parameter drift (IPD)” bezeichnet
 - Verursacht bspw. durch “Veralten” der Testaufgaben (wenn dort etwa von “Handys” die Rede ist, die heutige Grundschulkinder kaum noch kennen)



Längsschnittliche Analysen mithilfe von plausible values



- Was sind plausible values?

- die AV ist in IRT-Modellen latent, also unbeobachtet
- Mithilfe der beobachteten Werte (item responses) kann für jede Person eine hypothetische Fähigkeit modelliert werden
- Dieser personenspezifische Fähigkeitswert besitzt eine statistische Unsicherheit, es handelt sich also nicht um einen konkreten Wert, sondern eher eine Verteilung „plausibler“ Werte
- Zufallsziehungen aus dieser Verteilung bezeichnet man als plausible values (PVs). Sie können wie Imputationen unbeobachtete Werte verstanden werden, die man mithilfe beobachteter Werte möglichst genau schätzt
- Formal: PVs sind random draws aus der posterior distribution
 - Posterior als das Produkt der *a priori distribution* und der *likelihood*
 - A priori: was wissen wir vorab (also ohne zu testen) bereits über das zu messende Merkmal?
 - likelihood: was sagt uns der Test über das zu messende Merkmal?
- Beispiel Deutschtest: (a priori) bekannt ist, dass sich Jungen und Mädchen in ihrer mittleren Lesefähigkeit unterscheiden. Ebenso ist bekannt, dass die Lesefähigkeit mit dem sozio-ökonomischen Status im Elternhaus zusammenhängt
 - Das bedeutet aber: um diese plausible values verlässlich schätzen/ziehen zu können, müssen diese „a priori“-Merkmale (Geschlecht, SES) für die Testteilnehmer erhoben sein



Vor- und Nachteile von plausible values im Vergleich zu GLMMs



- **Nachteile**

- Aufwändiger: erfordert mehrere Analyseschritte
 1. Ziehen der Plausible values
 2. Rechnen mit Plausible values
- PVs nur geeignet für Modelle, die ausschließlich Prädiktoren auf Personenebene betrachten
- PVs sind imputiert, pro Person und AV gibt es mehrere imputierte Werte. Die Analysen müssen für jede Imputation durchgeführt und anschließend gepoolt werden

- **Vorteile**

- Liegen die plausible values erstmal vor, vereinfachen sich viele Analysen
 - Man braucht keine spezifische IRT-Software mehr, sondern kann Standardprogramme bzw. –analysen nutzen
 - In großen Datensätzen ist die Rechenzeit erheblich kürzer
- GLMMs funktionieren nur in der „Rasch“-Welt (1pl, dichotom), plausible values können auch für deutlich komplexere Modelle (2PL, 3PL, mixed IRT) gezogen werden

- **[persönliches] Fazit**

- PVs nehmen, wenn es nicht anders geht
- Viele große Schulleistungsstudien (z.B. PISA, IQB) stellen PV-Datensätze für Sekundäranalysen bereit