

# Personenfit in Veränderungsmodellen

Dipl.-Stat. Stefan Klein

Dipl.-Psych. Andreas Frey

Dipl.-Math. Ulrich Gauger

Institut für Psychologie  
Humboldt-Universität zu Berlin

# Veränderungsmessung in der klassischen Psychometrie

- Klassische (psychologische) Testtheorie (= KTT): Veränderungsmessung mittels Differenzen:

$$X_v^{Diff} = X_{v2} - X_{v1}$$

- Veränderungen sollen für jede Person einzeln gemessen werden

# Veränderungsmessung mit Differenzenscores: Problematik des Ansatzes

- Differenzenbildung führt zu hoher Fehlervarianz und somit geringer Reliabilität der Messungen
- Ausgangswertabhängigkeit: bei Extremgruppen wird die Veränderung durch Differenzen nur schlecht modelliert.
- Folgerung: die klassische Psychometrie rät von der Veränderungsmessung mit Differenzenscores ab
- Alternative Ansätze: Strukturgleichungsmodelle, Latent-Trait-Modelle, Markovmodelle

# Veränderungsmessung mit Latent- Trait-Modellen

- Veränderungsmessung für nominale / ordinale Variable
- Veränderung wird mit Hilfe einer latenten Variablen modelliert
- Bei dichotomen Items: Linear Logistic Test Model (= LLTM)
  - Modellierung der Antwortwahrscheinlichkeit einer Person  $\mathbf{v}$  für Item  $\mathbf{i}$  bei Zeitpunkt  $\mathbf{t}$ :

$$P(X_{vit} = x_{vit}) = \frac{\exp[x_{vit} (\theta_v - \beta_i + \delta_{gt})]}{1 + \exp[x_{vit} (\theta_v - \beta_i + \delta_{gt})]}$$

- Veränderungsmessung: nur für Gruppen von Personen möglich

# Schätzung für das LLTM

- Normale ML-Methode: führt zu nicht konsistenten Schätzungen der Itemparameter
- Conditional Maximum Likelihood-Methode:
  - Verwendung der bedingten Likelihood unter der Bedingung gegebener Summenscores
  - Führt zu konsistenten Schätzungen für Item- und Veränderungsparameter
  - Nachteil: nicht mit Standardsoftware wie PROC CATMOD oder PROC GENMOD durchzuführen

# Schätzung für das LLTM

- Alternative Möglichkeit: Schätzung eines zum LLTM äquivalenten loglinearen Modells mit PROC CATMOD

$$\log[P(X_{vit} = x_{vit})] = -\sum_{v,t} x_{vit} \beta_i + \sum_t \sum_g \delta_{gt} + \left( \sum x_{vit} u_r \right)$$

- Konsistente Schätzung der Veränderungsparameter möglich
- Nachteil: Schätzmethode setzt eine hochdimensionale Kontingenztafel voraus (z.B.: bei 7 Items und 2 Messzeitpunkten:  $2^{14}$  Zellen in der Kontingenztafel)

# Schätzung eines LLTM mit PROC CATMOD

- Syntax:

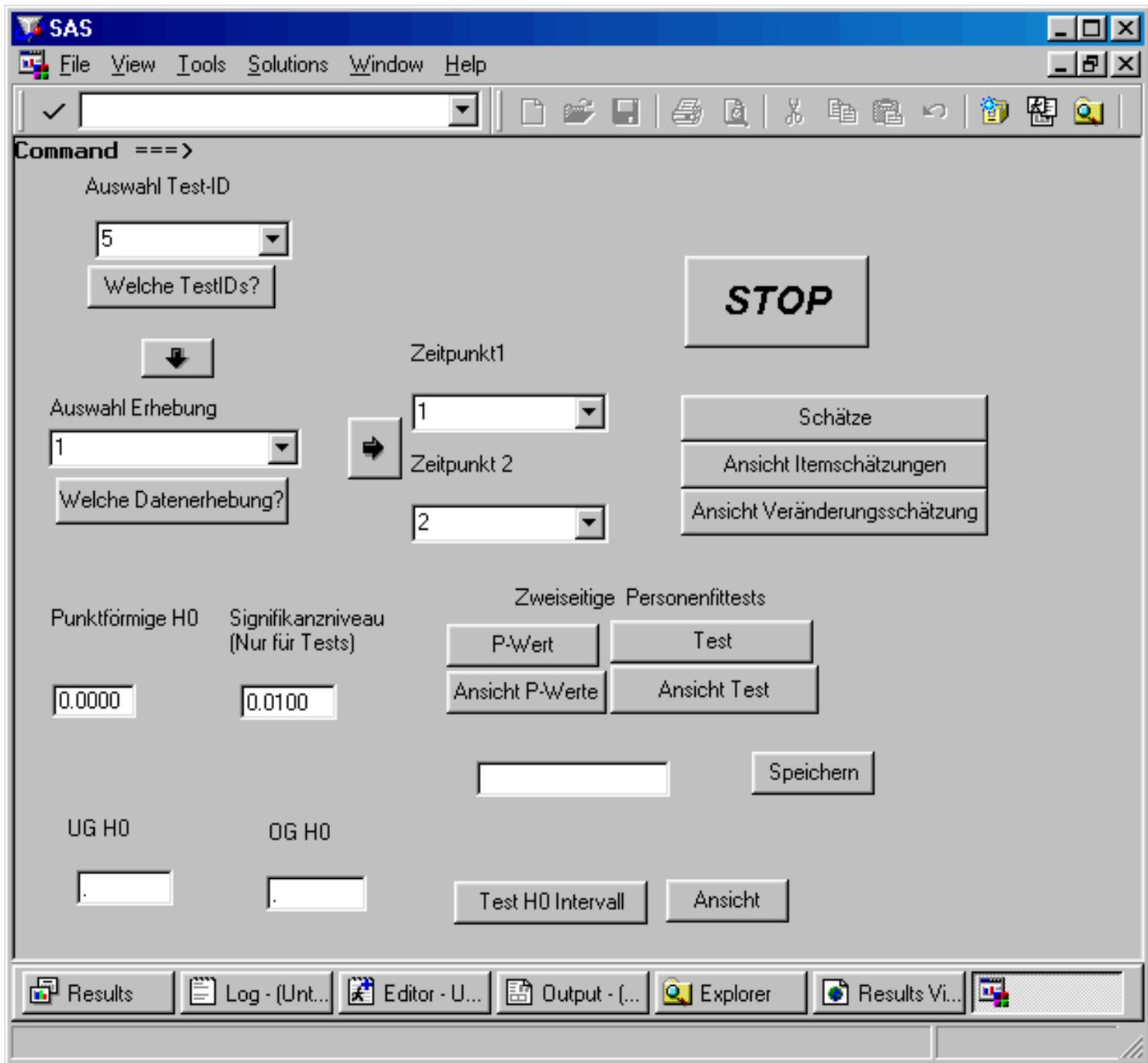
```
proc catmod data=big8herunter;
```

```
model antwortmuster=
```

```
    (1 0 ... 0    0... 0    1 ...    0,  
     0 1 ... 0    0... 1    0 1 ...    0,  
                                     ...  
     1 1 ... 1    7... 7    -1 ...    -1 );
```

# SAS-Programm für die Veränderungsmessung bei kategorialen Daten

- Schätzung von LLTMs mittels SAS/IML-Programm und Conditional Maximum Likelihood-Schätzung
- Zugrundeliegende Datenbank speichert wichtigste Informationen über den zu analysierenden Datensatz: Informationen über Versuchspersonen, Items, Datensätze
- Menügesteuerte Oberfläche
- Personenfittests, Likelihood-Quotiententests zur Bestimmung der Modellgüte
- Noch in Konstruktion: verbesserte Editiermöglichkeiten, weitergehende Schätz- und Testfunktionen



# Überprüfung der Anpassungsgüte eines Latent-Trait-Modells mittels Personenfitmaßen

- Personenfitmaße überprüfen, ob das Antwortmuster einer bestimmten Person mit den Voraussetzungen des gewählten Modells übereinstimmt
- Personenfitmaße bauen meist auf die Likelihood eines Antwortmusters auf.
- Antwortmuster mit niedriger Likelihood weisen eine schlechte Anpassung an das gewählte Modell auf.
- Absicherung der Entscheidung durch Signifikanztests

# Optimale Personenfitmaße für die Veränderungsmessung

- Bisher implementiert: Überprüfung vorgegebener Werte für den Veränderungsparameter
- Gleichmäßig beste unverfälschte Tests auf Personenniveau möglich
- Kontrolle des Gesamtsignifikanzniveaus:
  - Bonferroni
  - Methoden aus der Metaanalyse, z.B. „Adding of Logs“

# Anwendungsbeispiel: EORTC-QLQ-C30

- Fragebogen zur Lebensqualität bei Krebspatienten
- Insgesamt 30 Items und 9 Subskalen zur Messung des Konstrukts Lebensqualität
- Hier: Zusammenfassung der hoch korrelierenden Subskalen „Physische Beeinträchtigungen“ und „Rollenverhalten“
- Modellierung eines LLTMs für 7 Items und 2 Messzeitpunkte
- Nachträgliche Dichotomisierung der ursprünglich vierstufigen Items
- Gruppenbildung: Einteilung in ambulante/stationäre Patienten

# Ergebnisse der Modellschätzung

- Itemschwierigkeiten: liegen zwischen -3.3 und 2.7, d.h. breite Streuung der Itemschwierigkeiten
- Veränderungen: Ambulante Patienten weisen eine Veränderung von -0.13, stationäre eine Veränderung von 0.08 auf.
- Likelihood-Ratio-Tests:
  - Nullhypothese keine Veränderung: TG=0.89;  
 $\chi^2(2,0.89)=0.35$
  - Nullhypothese: Keine Auswirkung des Modells:  
TG=867.0;  $\chi^2(8,867.0)<0.0001$

# Ergebnisse der randomisierten Personenfittests

- Nullhypothese: keine Veränderung
- Getestet an allen 221 Personen der Stichprobe für Signifikanzniveau  $\alpha=0.1$
- Ergebnis: Bei 3 Personen kann die Nullhypothese abgelehnt werden, bei 7 weiteren Personen kann die Nullhypothese des randomisierten Tests mit Wahrscheinlichkeit 0.8 abgelehnt werden.
- Folgerung:

Der entdeckte schlechte Personenfit bei 10 Personen ist wohl zufällig bedingt.

# Zusammenfassung des Beispiels

Personenfittests bieten eine interessante Ergänzung zu Likelihood Ratio-Tests für Latent-Trait-Methode:

- Möglichkeit des Aussortierens von Probanden, bei denen das Modell nicht zutrifft
- Zielgerichtete Suche nach Variablen, mit denen die schlechte Modellanpassung einer Person zu erklären ist