
2.4 Indexbildung und Einblick in Skalierungsverfahren

2.4.1 Indexbildung

- Zusammenfassung mehrerer Indikatoren eines Konstrukts zu einer Kennzahl.
- Oft sind Konstrukte mehrdimensional
 - * Werden die Dimensionen jeweils durch einen Indikator erfasst?
 - * Welche Dimensionen fließen ein?
 - * Wie werden sie kombiniert?
 - * Jeder Ausprägungskombination wird ein Wert zugeordnet
 - * Zuordnungsregel: additiver Index
 - * Zuordnungsregel: multiplikativer Index

-
- * Zuordnungsregel: gewichteter Index $\sum_{j=1}^k g_j \cdot T_j$, wobei T_j mit $j = 1, \dots, k$ die Items bzw. Indikatoren widerspiegeln und g_j die (festgesetzten) Gewichte darstellen
 - Die Verknüpfungen durch Rechenoperationen als solche setzen ein höheres Skaleniveau voraus. Daher Verwendung von Indizes in vielen Situationen kritisch zu hinterfragen.
 - Index kontinuierlicher Variablen

2.4.2 Skalierungsverfahren I: Notation und probabilistische Grundstruktur

Methoden zur Konstruktion von Messinstrumenten:

Ermittle individuelle Ausprägungen einer latenten Eigenschaft mit Hilfe einer 'Batterie' von Fragen ('Items')

Notation

- Vorstellung: Universum von
 - * Personen und von
 - * Stimuli, Items, Aufgaben, etc.
- Personenfähigkeit, Einstellungsgrad, etc. latentes Merkmal Θ
- Ebenso: Schwierigkeit der Items _{i} : latentes Merkmal B (hier: 'großes β ')
- Ziehe zufällige Stichprobe von Personen $i, i = 1, \dots, n$, und von Items $j, j = 1, \dots, k$,
(Vorsicht: Oft werden Items auch mit i indiziert und Personen mit j)
- Θ_i : Personenfähigkeit, Einstellungsgrad etc. der i -ten Person in der Stichprobe,
Ausprägung ϑ_i

- B_j : Itemschwierigkeit des j -ten Items, Ausprägung β_j
- Datenmatrix $(T_{ij})_{i=1,\dots,n;j=1,\dots,k}$; T_{ij} als Antwort der Person i auf Frage/Item j
- Zur Ermittlung von $(\vartheta_i)_{i=1,\dots,n}$ und $(\beta_j)_{j=1,\dots,k}$ sehe diese Größen als Parameter eines statistischen Modells für die Daten $(T_{ij})_{i=1,\dots,n;j=1,\dots,k}$, d. h. modelliere Wahrscheinlichkeitsverteilung von T_{ij} als Funktion von ϑ_i und β_j

$$\mathcal{L}(T_{ij}) = h(\vartheta_i, \beta_j), \quad i = 1, \dots, n; j = 1, \dots, k$$

- Annahmen: Eindimensionalität, Aussagen zu Statements hängen nur mit der einen zu messenden Eigenschaft und Aufgabenschwierigkeit zusammen.
- Es macht Sinn, von einer probabilistischen Testtheorie zu sprechen, wobei dieser Begriff nur für kategoriale T_{ij} gebräuchlich ist.
- Typischerweise wird 'lokale stochastische Unabhängigkeit' vorausgesetzt:
Gegeben Θ_i und B_j , $i = 1, \dots, n; j = 1, \dots, k$, sind die einzelnen $(T_{ij})_{i=1,\dots,n;j=1,\dots,k}$

stochastisch unabhängig.

2.4.3 Skalierungsverfahren II: Metrische Beobachtungen

Ist $(T_{ij})_{i=1,\dots,n;j=1,\dots,k}$ metrisch, so ist ein natürlicher Ansatz in Verallgemeinerung des Grundmodells der klassischen Testtheorie der lineare Ansatz

$$\mathbb{E}(T_{ij}|\Theta_i = \vartheta_i, B_j = \beta_j) = \vartheta_i - \beta_j.$$

Für Kleinste-Quadrate-Methode vorstellen:

$$T_{ij} = \sum_{\ell=1}^n \vartheta_{\ell} X_{\ell,i} + \sum_{p=1}^k \beta_p X_{n+p,j}$$

für die $\tilde{n} = n \cdot k$ Beobachtungen mit

$$X_{\ell,i} = \begin{cases} 1 & \ell = i \\ 0 & \text{sonst} \end{cases} \quad 1 \leq \ell \leq n \quad \text{und} \quad X_{n+p,j} = \begin{cases} 1 & p = j \\ 0 & \text{sonst} \end{cases} \quad 1 \leq p \leq k.$$

Momentenmethode:

2.4.4 Skalierungsverfahren III: Binäre Beobachtungen

Situation T_{ij} binär, $T_{ij} = 1$ [bzw. $T_{ij} = 0$], Person i hat Aufgabe j richtig [bzw. falsch] beantwortet, oder Aussage j zugestimmt [abgelehnt]. Modelliert wird $p_1(T_{ij} = t_{ij} | \Theta_i = \vartheta_i, B_j = \beta_j)$. Wegen der lokalen stochastischen Unabhängigkeit gilt mit:

$$\mathbf{T} = (T_{ij}), i = 1, \dots, n; j = 1, \dots, k,$$

$$(\mathbf{\Theta}, \mathbf{B}) = (\Theta_i, B_j), i = 1, \dots, n; j = 1, \dots, k,$$

und entsprechenden Ausdrücken für die Ausprägungen

$$p(\mathbf{T} = \mathbf{t} | \mathbf{\Theta} = \vartheta, \mathbf{B} = \beta) = \prod_{i=1}^n \prod_{j=1}^k p(T_{ij} = t_{ij} | \Theta_i, B_j)$$

Itemcharakteristische Funktionen für binäre Items

Neuer Aspekt: Vergleich zwischen Subjekt und Stimuli (bzw. Item). Die Veranschaulichung erfolgt mithilfe sogenannter ICCs (*item characteristic curve*, traceline oder Itemcharakteristik);

-
- Die Grundform der ICCs wird in einigen Verfahren als bekannt (z.B. logistische Kurve im Rasch-Modell (s. u.)) angenommen.
 - Zeichne $p(T_{ij} = t_{ij} | \Theta_i = \vartheta_i, B_j = \beta_j)$ als Funktion von ϑ_i für verschiedene Werte von β_j , d. h.

Ausprägung der latenten Variablen wird auf der Abszisse abgetragen.

Lösungswahrscheinlichkeit bzw. Zustimmungswahrscheinlichkeit auf der Ordinate

- Interpretation monotoner Verlauf: Mit steigendem Wert der latenten Eigenschaft steigt Lösungs- oder Zustimmungswahrscheinlichkeit.
 - * Je fähiger die Person, desto höher Lösungswahrscheinlichkeit,
 - * bzw. je extremer die Einstellung der Person, desto höher Zustimmungswahrscheinlichkeit.

-
- Trennschärfe erkennt man an der Steigung der ICCs (verlaufen die Items parallel, so haben sie die gleiche Trennschärfe.)
 - Die sog. Item Response Theorie (IRT) umfasst viele Skalierungsverfahren, das wohl bekannteste ist das Rasch-Modell.

Dort setzt man:

$$P(T_{ij} = t_{ij} | \Theta_i = \vartheta_i, B_j = \beta_j) = \frac{\exp[(\theta_i - \beta_j) \cdot t_{ij}]}{1 + \exp(\theta_i - \beta_j)}$$

Auch in der PISA Studie werden das Rasch-Modell und einige Erweiterungen verwendet, um die Vorbereitung von 15-jährigen Schülern auf die Wissensgesellschaft zu messen.

Vorgehen:

1. Schätzung der Aufgabenschwierigkeiten
2. Schätzung der Personenparameter

-
- Nichtstandard-Fall im Maximum-Likelihood-Kontext
 - Zur Schätzung der Parameter kann z.B. die sog. bedingte Maximum-Likelihood Schätzung oder die sog. marginale Likelihood-Schätzung verwendet werden.
 - Wenn das Rasch-Modell gilt, entsprechen die Personenparameter (mindestens) einer Intervallskala.
 - Wenn das Rasch-Modell gilt, dann ist die geschätzte Fähigkeit oder Einstellung eine monotone Transformation der Summe der gelösten Items. Diese bilden dann eine ordinale Skala.
 - Zentraler Vorteil der IRT Modelle ist, dass die Annahmen explizit angegeben werden können und Tests zur empirischen Überprüfung der Annahmen existieren.

Weitere Skalierungsverfahren

- Mehrdimensionale Rasch-Modelle
- Birnbaum Modell (auch 2 PL-Modell genannt)
- 3 PL-Modell
- Partial-Credit Modell, Rating-Scale-Modell, Graded-Response Modell
- Im weiteren Sinn auch Paarvergleichmodellierung :

2.4.5 Modellabweichungen, Itembereinigung

- 'Differential Item Functioning' ; Aufgabenschwierigkeit hängt von externen Faktoren ab, eine Aufgabe ist also nicht gleich schwer für alle Teilnehmer
- Manche Items nicht 'trennscharf', kein/geringer Einfluss von individueller Fähigkeit

Likert-Skala für die Situation aus 2.4.3

- Likert-Skala bei 'als metrisch angenommen Größen: Statements von „stimme voll zu“, „stimme eher zu“, „teils/teils“, „stimme eher nicht zu“, „stimme nicht zu“
- übliches Vorgehen:
 - * „Ratings“ werden aufsummiert (▷ Intervallskala vorausgesetzt!?) $\hat{\vartheta} =$
 - * Prüfung der Trennschärfe (wie gut werden hohe Eigenschaftswerte und niedrige Eigenschaftswerte durch die Kategorien separiert) und der Reliabilität (Cronbachs α , s. u.)
 - * ausreichend reliable und trennscharfe Items bleiben, aus diesen wird Summe oder Mittelwert gebildet

- Probleme

- * Eindimensionalität häufig verletzt
- * kein Nachweis eines Skalenniveaus
- * abhängig von der Stichprobe
- * Verfahren ist in gewisser Weise zirkulär.

Exkurs: Einige Regeln zur Formulierung von Items

- eindimensionale Items
- Frage nach gegenwärtigem Zustand
- keine Suggestion durch Tatsachenbeschreibungen,
- Items sollten den Wertebereich, in dem die Befragten liegen, abdecken
- einfache, klare, kurze, verständliche Struktur, keine Mehrdeutigkeit, keine doppelte Verneinung
- insbesondere natürlich widerspruchsfrei.
- Antwortkategorien erschöpfend