

2.3 Das Grundmodell der klassischen Testtheorie (KTT)

Wie kann man die Messproblematik statistisch fassen?

Schnittpunkt zwischen üblicher empirischer Methodik und fortgeschrittenerer statistischer Methodik.

- Man könnte ähnlich wie beim Inferenzfehler vorgehen und sagen: Wenn der Messfehler schon prinzipiell unvermeidlich ist, so sollte man ihn wenigstens statistisch kontrollieren/ mitberücksichtigen.
- Betrachtet wird hier ein Modell für die Messung einer metrischen, potentiell mehrdimensionalen, latenten Größe.
- Verallgemeinerungen bezüglich der Messsituation bzw. der expliziten Mitberücksichtigung des Messprozess in statistischen Modellen werden später nochmals aufgegriffen (siehe die Abschnitte „Skalierungsverfahren“ (Kap. 2.5) und „Fehler-in-den-Variablen-Modellen“ (in Kap. 7)

- Notation: Für jeden Vektor $a \in \mathbb{R}^r$ bezeichne $a[\ell]$ die ℓ -te Komponente, also $a = (a[1], a[2], \dots, a[r])^T$.

2.3.1 Im Kontext einer einzelnen Messung

$$\begin{array}{ccccc} C & = & \Gamma & + & \Delta & (2.1) \\ \uparrow & & \uparrow & & \uparrow & \\ \text{gemessener} & & \text{wahrer} & & \text{Messfehler} & \\ \text{Wert} & & \text{Wert} & & & \end{array}$$

Da sowohl Γ als auch Δ latent sind, ist diese Gleichung *immer* erfüllt, man setze einfach $\Delta := T - \Gamma$. Entscheidend sind daher, die Zusatzannahmen, die man unterlegt. Diese sind in konkreten Anwendung unbedingt kritisch zu hinterfragen.

Was bedeutet hier eigentlich “Modell“?

Def. 2.1.

Sei $(\mathbf{C}, \mathbf{\Gamma}, \mathbf{\Delta})$ ein Tripel von metrischen mehrdimensionalen Merkmalen und C, Γ, Δ die daraus durch reine Zufallsauswahl eines Elements gewonnenen dreidimensionalen Zufallsvektoren.

1. $(\mathbf{C}, \mathbf{\Gamma}, \mathbf{\Delta})$ und (C, Γ, Δ) mit

$$C = \Gamma + \Delta \quad (2.2)$$

soll als dem *unzentrierten Grundmodell der klassischen Testtheorie* gehorchend bezeichnet werden, wenn gilt:

- (a) $\mathbb{E}(\Delta)$ und $Cov(\Delta)$ existieren.¹³
- (b) Γ und Δ sind komponentenweise unabhängig: Für alle $\ell_1 = 1 \dots, r$ und $\ell_2 = 1 \dots, r$ gelte $\Gamma[\ell_1]$ und $\Delta[\ell_2]$ sind stochastisch unabhängig
- (c) Messfehler bei verschiedenen Komponenten sind voneinander unabhängig: $\Delta[1], \Delta[2], \dots, \Delta[r]$ sind stochastisch unabhängig
2. Gilt zusätzlich, dass kein systematischer Messfehler vorliegt, d.h.

$$\mathbb{E}(\Delta) = 0,$$

so sei vom *(zentrierten) Grundmodell der klassischen Testtheorie* gesprochen.

¹³Diese Bedingung wird in der Literatur meist implizit vorausgesetzt.

Bem. 2.2.

- Werden einzelne Komponenten $\Gamma[\tilde{\ell}]$ von $\vec{\Gamma}$ messfehlerfrei erhoben, so wird einfach $\Delta[\tilde{\ell}] = 0$ gesetzt. Man kann das Modell erweitern, wenn man vorausgesetzt, dass alle nicht metrischen Variablen exakt gemessen werden, der Ansatz (2.1) passt also auch für sie.
- Man müsste eigentlich noch eine *starke* und eine *schwache* Form unterscheiden. Die hier gegebene Formulierung entspräche dann der starken Form; bei der schwachen Form wird nur jeweils Unkorreliertheit statt Unabhängigkeit gefordert.
- Vor allem in der Psychometrie wird allerdings häufig implizit zusätzlich vorausgesetzt, dass $(C, \Gamma, \Delta)^T$ multivariat normalverteilt ist, wodurch starke und schwache Form zusammenfallen.

Def. 2.3.

In der in Definition 2.1 eingeführten Situationen heie dann jeweils *C* *Messung des Konstrukts* Γ (mit *Messfehler* Δ) im Sinne des unzentrierten bzw. zentrierten Grundmodells der klassischen Testtheorie.

Erste Konsequenzen

2.3.2 Individuelle Messungen, mehrere Personen

Jetzt Stichprobe von Umfang n

Def. 2.4.

Die in Definition 2.1 eingeführten Bezeichnungen seien auf *i.i.d* Stichproben $(C_i, \Gamma_i, \Delta_i)$, $i = 1, \dots, n$, des Tripels $(\mathbf{C}, \mathbf{\Gamma}, \mathbf{\Delta})$ von Merkmalen ausgedehnt, wenn $(C_i, \Gamma_i, \Delta_i)$ für alle i die entsprechenden Bedingungen erfüllt.

2.3.3 Wiederholungsmessungen

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

$C^{(j)}$, j -te Wiederholung der Messung von Γ_i , analog $C_i^{(j)}$

Def. 2.5.

In analoger Weise seien die Bezeichnungen auch auf Wiederholungsmessungen ausgedehnt, wenn zusätzlich gilt

- a) im Fall von $(C^{(j)}, \Gamma, \Delta^{(j)})$, $j = 1, \dots, p$:
- a1) Für jedes j erfüllt das Tripel $(C^{(j)}, \Gamma, \Delta^{(j)})$ die entsprechenden Bedingungen.
 - a2) $(\Delta^{(1)}, \dots, \Delta^{(j)}, \dots, \Delta^{(p)})$ sind identisch verteilt und $(\Gamma, \Delta^{(1)}, \dots, \Delta^{(j)}, \dots, \Delta^{(p)})$ sind gemeinsam unabhängig.
- b) im Fall von $(C_i^{(j)}, \Gamma_i, \Delta_i^{(j)})$, $i = 1, \dots, n$, $j = 1, \dots, p$:
- b1) Für jedes feste i , $i = 1, \dots, n$, genügen die Tripel $(C_i^{(j)}, \Gamma_i, \Delta_i^{(j)})$, $j = 1, \dots, p$, dem entsprechenden Grundmodell im Sinne von Teil a).
 - b2) Für jede beliebige Auswahl $(j_1, \dots, j_i, \dots, j_n)$ genügen die Tripel $(C_i^{(j_i)}, \Gamma_i, \Delta_i^{(j_i)})$, $i = 1, \dots, n$, dem entsprechenden Grundmodell im Sinne von Def. 2.1.

d) Einige weitere Konsequenzen

Bem. 2.6.

Gehorcht $(C_i^{(j)}, \Gamma_i, \Delta_i^{(j)})$, $i = 1, \dots, n$, $j = 1, \dots, p$, dem unzentrierten (bzw. zentrierten) Grundmodell der klassischen Testtheorie, so kann man zeigen:

a) Für jedes i genügt

$$\bar{C}_i := \frac{1}{p} \underbrace{\sum_{j=1}^p C_i^{(j)}}_{\text{Danke!}}$$

dem unzentrierten (bzw. zentrierten) Grundmodell der klassischen Testtheorie.

b) Setzt man die starke Form voraus, so gilt die sogenannte *lokale stochastische Unabhängigkeit*:

Für jedes i sind für jedes $j_1 \neq j_2$ die Messungen $C_i^{(j_1)}$ und $C_i^{(j_2)}$ bedingt unabhängig gegeben Γ_i , d.h. es gibt keine über Γ_i hinaus gehende Abhängigkeit zwischen $C_i^{(j_1)}$ und $C_i^{(j_2)}$.

2.4 Adäquation; Gütekriterien: Objektivität, Reliabilität, Validität

Eine Messung im schwachen Sinn kann zunächst relativ willkürlich und inhaltlich sinnlos sein.

Bei der Frage nach der Adäquation einer Formalisierung ist die Frage nach der Güte der Messungen ein zentraler Punkt. Bei der Beurteilung unterscheidet man typischerweise drei aufeinander aufbauende Aspekte:

i) Objektivität:

Grad der Unabhängigkeit der Messung von Einflüssen außerhalb untersuchten Einheit

* Durchführungsobjektivität:

* Auswertungsobjektivität:

* Interpretationsobjektivität:

ii) Reliabilität (Zuverlässigkeit)

In welchem Ausmaß führt eine wiederholte Messung zu demselben Ergebnis?

iii) Validität (Gültigkeit)

Grad der Genauigkeit, mit der ein Verfahren oder eine Messung das misst, was es messen soll

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

Es sei im Folgenden festgesetzt, dass Γ eindimensional sei mit $0 < \mathbb{V}(\Gamma) < \infty$.

2.4.1 Beurteilung der Reliabilität

Def. 2.7. *Maßzahl für die Reliabilität (theoretische Reliabilität):*

Gegeben sei eine Messung C des eindimensionalen Konstrukts Γ (mit Messfehler Δ) im Sinne des unzentrierten Grundmodells der klassischen Testtheorie. Dann heißt

$$Rel(\Gamma, C) := \frac{Var(\Gamma)}{Var(C)}$$

theoretische Reliabilität von C für Γ .¹⁴

Bem. 2.8. *Rolle der Messfehlervarianz*

¹⁴Wegen der Voraussetzung $Var(\Gamma) > 0$ ist auch $Var(C) > 0$, und damit ist $Rel(\Gamma, C)$ wohldefiniert.

Satz 2.9.

In der Situation von Def. 2.7 gilt für die Korrelation $\varrho(\Gamma, C)$ von Γ und C :

$$\varrho(\Gamma, C) = \sqrt{Rel(\Gamma, C)}$$

Beweis

Bem. 2.10.

Alle hier gegebenen Argumentationen stehen und fallen mit der Unkorreliertheit von Γ und Δ !!

„Schätzung“ der Reliabilität

Satz 2.11. [Hauptsatz der Reliabilitätsschätzung] Seien $C^{(1)}, C^{(2)}$ zwei Wiederholungsmessungen für Γ , die dem unzentrierten Grundmodell der klassischen Testtheorie im Sinne von Definition 2.5 genügen. Dann gilt für den Korrelationskoeffizienten $\rho(C^{(1)}, C^{(2)})$

$$\rho(C^{(1)}, C^{(2)}) = Rel(\Gamma, C^{(1)}) = Rel(\Gamma, C^{(2)}).$$

Beweis:

Satz 2.12. (*Spearman-Brown Formel*)

Seien $C^{(1)}, C^{(2)}, \dots, C^{(p)}$ Wiederholungsmessungen, die dem unzentrierte Grundmodell der klassischen Testtheorie im Sinne von Definition 2.5 genügen. Dann gilt für

$$\bar{C} := \frac{1}{p} \sum_{j=1}^p C^{(j)} :$$

- i) \bar{C} ist eine Messung des Konstrukts Γ im Sinne des unzentrierten Grundmodells der Klassischen Testtheorie (vgl. Bem 2.6).
- ii) Es gilt für jedes $j = 1, \dots, p$

$$Rel(\Gamma, \bar{C}) = \frac{p \cdot Rel(\Gamma, C^{(j)})}{1 + (p - 1)Rel(\Gamma, C^{(j)})}$$

Beweis:

Praktische Umsetzung

- *Validierungsdaten:*
- Test-Retest-Methode:
- Paralleltestmethode:
- Split-Half-Methoden zur Erzeugung paralleler Tests:
- Cronbachs Alpha (berücksichtigt die verschiedenen Aufteilungsmöglichkeiten bei p Messungen).

Normierte Form:

$$\alpha := \frac{p \cdot \bar{\varrho}}{1 + \bar{\varrho}(p - 1)}$$

mit $\bar{\varrho}$ als durchschnittlicher Korrelationskoeffizient zwischen zwei Messungen

Bsp. 2.13. *Man betrachte die in Abb. 2.1 gegebene Situation und bestimme Cronbachs α*

Reliabilitätsstatistiken

| Cronbachs Alpha | Cronbachs Alpha für standardisierte Items | Anzahl der Items |
|-----------------|---|------------------|
| ,799 | ,801 | 4 |

Inter-Item-Korrelationsmatrix

| | FRAU, LIEBER MANN BEI D. KARRIERE HELFEN? | FRAU, NICHT ARBEITEN BEI KLEINKIND? | FRAU, ZU HAUSE BLEIBEN+KINDER VERSORGEN ? | FRAU, NACH HEIRAT ARBEITSPL. FREIMACHEN ? |
|---|---|-------------------------------------|---|---|
| FRAU, LIEBER MANN BEI D. KARRIERE HELFEN? | 1,000 | ,385 | ,547 | ,457 |
| FRAU, NICHT ARBEITEN BEI KLEINKIND? | ,385 | 1,000 | ,583 | ,446 |
| FRAU, ZU HAUSE BLEIBEN+KINDER VERSORGEN? | ,547 | ,583 | 1,000 | ,587 |
| FRAU, NACH HEIRAT ARBEITSPL. FREIMACHEN? | ,457 | ,446 | ,587 | 1,000 |

Abbildung 1: Beispiel aus dem ALLBUS 2008, bereitgestellt von GESIS, Leibniz-Institut für Sozialwissenschaften.

2.4.2 Beurteilung der Validität

Def. 2.14. (*Theoretische Validität:*)

Sei C eine Messung für Γ , die dem zentrierten Grundmodell der Klassischen Testtheorie genügt. Dann heißt $\rho(C, \Gamma)$ theoretische Validität.

Bem. 2.15.

Im Lichte von Satz 2.9 entspricht also die Validitätsmessung der Reliabilitätsmessung, wobei aber zusätzlich gefordert wird,

$$\mathbb{E}(\Delta) = 0,$$

also die zentrierte Variante vorliegt.

- „Axiomatisierung“

- *Inhaltsvalidität*: Alle Dimensionen des Konstrukts erfasst – und nur diese!

Goodhart's Law:

Praktische Beurteilung der Validität:

- Optimal wären wieder *Validierungsdaten*:
- *Expertenbefragung*
- *Kriteriumsvalidität*: Hoher Zusammenhang zwischen Messwerten und einem anderen gemessenem Kriterium („externes Kriterium“)
- * Prädiktive Validität:
- * Konkurrente Validität:
- *Konstruktvalidität*:
- * Kriterien:
Konvergenz:

Diskriminanz:

* Multitrait-Multi-Method-Matrix

2.4.3 Mehrere Merkmale, Korrelations- und Regressionsanalyse

Satz 2.16.

Seien C und T Messungen für Γ bzw. Θ mit Messfehler Δ bzw. Ψ im Sinne des Grundmodells der klassischen Testtheorie und Δ und ψ stochastisch unabhängig. Dann gilt:

$$\varrho(\Gamma, \Theta) = \frac{\varrho(C, T)}{\sqrt{\text{Rel}(\Gamma, C)} \cdot \sqrt{\text{Rel}(\Theta, T)}} \quad (2.3)$$

Beweis:

Bem. 2.17.

Die rechte Seite wird oft als Masszahl der Kriteriumsvalidität verwendet. Sind die Voraussetzungen des Satzes erfüllt, so lässt sich folgern:

1. Unter Messfehlern ist $\text{Rel}(\Gamma, C) < 1$ und/oder $\text{Rel}(\Theta, T) < 1$; also zeigen sich die wahren Zusammenhänge *abgeschwächt* im zugehörigen Indikator (engl. „attenuation“).
2. $\sqrt{\text{Rel}(\Gamma, C)} \cdot \sqrt{\text{Rel}(\Theta, T)}$ kann als Korrekturfaktor gesehen werden, um von $\varrho(C, T)$ auf $\varrho(\Gamma, \Theta)$ schliessen zu können („*correction for attenuation*“).
3. *Seien* $\beta(C, \Theta)$ bzw. $\beta(C, T)$ die „theoretischen Regressionskoeffizienten“ einer klassischen linearen Einfachregression mit Absolutglied, Γ bzw. C als abhängige Variable und Θ bzw. T als Kovariable. Dann gilt:

$$\beta(C, T) = \text{Rel}(\Theta, T) \cdot \beta(\Gamma, \Theta) \quad (2.4)$$

Beweis:

4. Die Herleitungen benutzen explizit, dass

$$\text{Cov}(\Gamma, \Psi) = \text{Cov}(\Delta, \Theta) = \text{Cov}(\Delta, \Psi) = 0.$$

Bestehen Abhängigkeiten, so gilt im Allgemeinen (2.3) nicht mehr. Die entsprechenden Korrekturfaktoren können grösser oder kleiner sein.

5. Die Aussagen gelten gleichermaßen für das unzentrierte Grundmodell der Klassischen Testtheorie, denn die Bedingungen $\mathbb{E}(\Delta) = 0$ und $\mathbb{E}(\Psi) = 0$ werden in der Herleitung nicht benötigt: Valide Messinstrumente und nicht valide Messinstrumente mit derselben Reliabilität liefern dieselbe Korrelation.
6. In 2.4 geht die Reliabilität $Rel(\Gamma, C)$ der Messung der abhängigen Variable nicht mit ein (in der Tat wird der entsprechende Messfehler von dem ϵ der Regression sozusagen aufgezogen).
7. Die in 2.4 beobachtete erneute Attenuierung setzt essentiell die Gültigkeit des unzentrierten Grundmodells der Klassischen Testtheorie voraus. Sie lässt sich auch

nicht auf lineare Regression mit mehreren Kovariablen oder auf beliebige lineare Regressionsmodelle verallgemeinern.¹⁵

¹⁵siehe z.B. Carroll, Ruppert, Stefanski, Crainiceanu (2006): Measurement Error in Nonlinear Models: A Modern Perspective, Chapman&Hall.