

Reliabilität

Bißantz, Jalynskij, Kupffer & Prestele

BF3 Testtheorie

Reliabilität als Konzept (20 min)

Einheit: Was ist Reliabilität?

Inhaltliche Schwerpunkte der Einheit:

- 1 Reliabilität als Konzept
- 2 Definition der Reliabilität
- 3 Das Reliabilitätsproblem

Ziel: Wiederholung der Konzepte (re-fresher) & Problematisierung

Mittel: Gruppenübung (15 min) & Input (5min)

Reliabilität als Konzept (20 min)

Gruppenübung: 6 Gruppen (a 3-4 Leute) Mosburger & Kelava, S. 307 - 309

- Reliabilität als Maß der Messgenauigkeit
- Reliabilität, Wahrer Wert und Messfehler
- Testwertvariable als Summe der Itemvariable
- Definition: Reliabilität einer Testvariable
- Definition: Reliabilität einer Itemvariable
- Reliabilität, Wertebereich und dessen Bedeutung
- Reliabilität und Objektivität
- Reliabilität und Validität

Posten der Ergebnisse im Olat-Forum

Auflösung: Reliabilität als Konzept

Messinstrument mit hoher Messgenauigkeit, Messergebnisse mit geringem Messfehler

Perfekte Reliabilität: Abwesenheit von zufälligem Messfehlern

$$E(\epsilon) \rightarrow 0 : E(X) = \tau$$

Testwert ist der Summenscore über alle Itemvariablen (\in Itemuniverse)

$$X = \sum_{i=1}^n X_i$$

Reliabilität einer Testwertvariablen (Y): $Rel(y) = \frac{Var(T)}{Var(X)} = \frac{Var(T)}{Var(T) + Var(E)}$

Reliabilität einer Itemvariable (Y): $Rel(x_i) = \frac{Var(\tau_i)}{Var(x_i)} = \frac{Var(\tau_i)}{Var(\tau_i) + Var(\epsilon_i)}$

$Rel \in [0, 1]$, wobei $Rel = 1$ Abwesenheit von Messfehlern; d.h. vollständig reliable Messung (vice versa)

Objektivität \rightarrow Reliabilität (via: Messbedingungen standardisieren)

Reliabilität \rightarrow Validität (v.a.: Beständigkeit gleicher Testergebnisse bei wiederholter Messung)

Das Reliabilitätsproblem

Problem: Wir kennen die True-Score- und Fehlervarianz nicht. Die Messwerte bei einer einzigen Messung sind lediglich *Schätzer* der wahren Werte, die *approximativ* dem wahren Wert entsprechen:

$$\tau = E(X) | E(\epsilon) = 0 \quad (1)$$

Damit lässt sich mit einer *Einzelmessung* die Reliabilität nicht eindeutig *bestimmen*! (siehe auch: Moosburger & Kelava, 2021 : 210). Wir müssen sie *schätzen*.

Methoden der Reliabilitätsschätzung (20 min)

Einheit: Methoden der Reliabilitätsschätzung

Inhaltliche Schwerpunkte der Einheit:

- ① Lösungsansatz zum Reliabilitätsproblem
- ① Retest-Reliabilität
- ② Paralleltest-Reliabilität
- ③ *Testhalbierungsreliabilität*
- ④ *Interne Konsistenz*

Ziel: Wiederholung und Vertiefung der Konzepte, Umsetzung in R

(?) Mittel: Input (10 min) & R-Übung (5min)

Lösungsansatz zum Reliabilitätsproblem

*“Aber auch ohne die wahren Werte einzelner Personen zu kennen, kann das Varianzverhältnis als Maß für die Messgenauigkeit geschätzt werden, wenn man die Ebene der einzelnen Personen und einzelnen Items verlässt und stattdessen **alle Items, aus denen sich ein Test zusammensetzt**, sowie die **Messungen mehrerer Personen betrachtet**: Wird ein latentes Merkmal anhand mehrerer Items gemessen, so liegen **Mehrfachmessungen desselben Merkmals mit unterschiedlichen aber ähnlichen Messinstrumenten/Items** vor, die **zu einer Testwertvariablen aufsummiert werden können**, sofern sie zumindest die Bedingung der Eindimensionalität¹ erfüllen.”* (Mosburger & Kelava, 2020: 310 – Hervorhebungen nicht im Original)

¹Die Bedingung der Eindimensionalität können und sollten Sie überprüfen (Hilsmittel: CFA). Die Unkorreliertheit der Messfehler ($\text{Cov}(\epsilon_i, \epsilon'_i) = 0$) ist dabei eine Basisvoraussetzung, für die Erfüllung der Bedingung. (Siehe: ebd., 14.2.2)

Lösungsansatz zum Reliabilitätsproblem

Lösungsansatz in a Nuthshell: Mehrfachmessungen

Reliabilität(-sschätzung) \Rightarrow Mehrfachmessungen! D.h. Alle Methoden zur Reliabilitätsschätzung setzen eine Mehrfachmessung des Konstruktes voraus!

Möglichkeiten zur Mehrfachmessung (Population/Itemuniversum)

- ① Wiederholte Messung anhand derselben/verschiedener Testdurchläufe²
- ② Verschiedene Items innerhalb eines Tests³

²Erinnerung: Sitzung 04-KTT (v.a. Übungsaufgaben 2 & 3)

³Erinnerung: "Item-Universum" & "Cronbach's Alpha"

Einflussfaktoren auf die Höhe der Reliabilität

Anwendung: Konfidenzintervalle

Francois, Romain. 2020. *Bibtex: Bibtex Parser*.

<https://github.com/romainfrancois/bibtex>.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

<https://www.R-project.org/>.

Revelle, William. 2021. *Psych: Procedures for Psychological, Psychometric, and Personality Research*.

<https://personality-project.org/r/psych/%0Ahttps://personality-project.org/r/psych-manual.pdf>.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC.

<https://bookdown.org/yihui/rmarkdown-cookbook>.