Reliabilität

Bißantz, Jalynskij, Kupffer & Prestele

BF3 Testtheorie

1. Reliabilität als Konzept (20 min)
2. Methoden der Reliabilitätsschätzung (30 min)
3. Einflussfaktoren auf die Höhe der Reliabilität
4. Anwendung: Konfidenzintervalle
5. Selbststudium

Section 1

Reliabilität als Konzept (20 min)

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, Wertebreich und dessen Bedeutung
- Reliabilität und Objektivittät
- 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
$E(\epsilon) \to 0 : E(X) = \tau$

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 Mesfehlern; d.h. vollständig reliable Messung (vice versa)

Objektivität -> Reliabilität (via: Messbedingungen standardisieren)

Reliabilität -> 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:

$$\tau = E(X)|E(\epsilon) = 0 \tag{1}$$

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

Section 2

Methoden der Reliabilitätsschätzung (30 min)

Methoden der Reliabilitätsschätzung (30 min)

Einheit: Methoden der Reliabilitätsschätzung

Inhaltliche Schwerpunkte der Einheit:

0. 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)

Lösunganstz 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 Eindimensionalität*[1] *erfüllen." (Mosburger & Kelava, 2020: 310 – Hervorherbungen nicht im Original)*

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

Lösunganansatz zum Reliabilitätsproblem

Lösungsansatz in a Nuthshell: Mehrfachmessungen

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

Möglichkeiten zur Mehrfachmessung (Population/Itemuniversum)

1. Wiederholte Messung anhand derselben/verschiedener Testdurchläufe[2]
   - Test/Test Reliabilität(en)
2. Verschiedene Items innerhalb eines Tests[3]
   - Interne Konsistenz (Cronbach's alpha)

[2] Erinnerung: Sitzung 04-KTT (v.a. Übungsaufgaben 2 & 3)
[3] Erinnerung: "Item-Universum" & "Cronbach's Alpha"

Übungsaufgabe 1: Selbstexperiment

Retest & Paralleltest-Reliabilität (Wiederholung)

Beim Restest wird der Test an der gleichen Stichprobe zu zwei verschiedenen Zeitpunkten durchgeführt. Die Restest-Reliabilität berechnet sich dann als Korrelation der Testwerte. Die Retest-Reliabilität setzt (a) konstante wahre Werte und (b) konstante Messfehlereinflüsse voraus. Unter diesem Annahmen entspricht die Korrelation der Testwerte dem Anteil der wahren Varianz an der Varianz der Testwerte.

Beim Paralleltest werden derselben Stichprobe parallele Testformen vorgegeben. Die Paralleltest-Reliabilität berechnet sich dann als Korrelation der resultierenden Testwerte. Bei welcher Testart wird dieses Verfahren häufig eingesetzt? Nun, viel spricht für den Leistungstest. Ein Problem mit parallelen Testformen sind Übungseffekte und Transfereffekte. Zur Überprüfung der Parallelität von Testformen setzem wir die konfirmatorische Faktorenanalyse ein.

Testhalbierungsreliabilität (Split-half)

Kochrezept: Testhalbierungsreliabilität

Zubereitungszeit: 5-10 min
Schwierigkeit: mittel

Zutaten:

- (Simultierter) Test voll mit Items
- Partitionierungsmethode[4]
- Korrelationskoeffizient
- Spearman-Brown

Den Test voll mit Items mit der Partitionierungsmethode in zwei parallele Testhälften aufteilen. Anschließend die Halbtestreliabilität mit der *Spearman-Brown-Korrektur* zur vollständigen Reliabilität aufwerten.

[4] zum Beispiel: *Odd-Even Aufteilung*, Zeitpartitionierungsmethode, Selektion von Itemzwillinge oder *Zufallsaufteilung*

Simulation: Test voll mit Items (..in R)

```r
# Tau-parallel
set.seed(123)
M <- 8
mu <- c(5,4,3,4,5,3,5,4)
# Kovarianzmatrix
Sigma <- matrix(
  c(.8, .5, .5, .5, .5, .5, .5, .5,
    .5, .8, .5, .5, .5, .5, .5, .5,
    .5, .5, .8, .5, .5, .5, .5, .5,
    .5, .5, .5, .8, .5, .5, .5, .5,
    .5, .5, .5, .5, .8, .5, .5, .5,
    .5, .5, .5, .5, .5, .8, .5, .5,
    .5, .5, .5, .5, .5, .5, .8, .5,
    .5, .5, .5, .5, .5, .5, .5, .8),
  M,M)
N <- 1e3
X <- data.frame(MASS::mvrnorm(N, mu, Sigma))
```

Partiotionierung & Halbtestreliabilität (..in R)

Partitionierungsmethode: Odds-Even Aufteilung

```
set.seed(123)
even <- seq(1,8, by=2)
uneven <- seq(2,8, by=2)
rsx_even <- rowSums(X[,even])
rsx_uneven <- rowSums(X[,uneven])
# Halbtestreliabilität
(Rel_halb <- cor(rsx_even, rsx_uneven))
```

```
## [1] 0.8710395
```

⇒ Mit der Halbtestreliabilität soll nun die vollständige Reliabilität geschätzt werden.

Vollständige Reliabilität (..händisch)

Example

$$
\begin{aligned}
Rel(X_{voll.}) &= \frac{2 \cdot Corr(X_p, X_q)}{1 + Corr(X_p, X_q)} \\
&= \frac{2 \cdot Rel_{halb}}{1 + Rel_{halb}} \\
&\approx \frac{2 \cdot 0.89}{1 + 0.89} \\
&\approx \frac{2 \cdot 0.89}{1 + 0.89} \\
&\approx \frac{1.78}{1.89} \\
&\approx \frac{1.78}{1.89} \\
&\approx 0.94
\end{aligned}
$$

Spearman-Brown Korrektur (..in R)

Spearman-Brown Formel

$$Rel(X_{voll.}) = \frac{2 \cdot Corr(X_p, X_q)}{1 + Corr(X_p, X_q)} = \frac{2 \cdot Rel_{halb}}{1 + Rel_{halb}} \tag{2}$$

…in R-isch:

```
Rel_SBK <- function(X_p, X_q) {
  rs_p <- rowSums(X_p)
  rs_q <- rowSums(X_q)
  r <- cor(rs_p, rs_q)
  2 * r / (1+r)
}
```

Vollständige Reliabilität (..in R)

Bei der Überprüfung der händischen Berechnung kommt nun `Rel_SBK()` zum Einsatz. Die Aufteilung ist nach wie vor "Odds-Even."

```
set.seed(123)
X <- data.frame(MASS::mvrnorm(N, mu, Sigma))
even <- seq(1,8, by=2)
uneven <- seq(2,8, by=2)
X_p <- X[, even]
X_q <- X[, uneven]
Rel_SBK(X_p, X_q)
```

```
## [1] 0.9310755
```

Übungsaufgabe 2: Selbsttest

Example

Versuchen Sie es nun selbst! Sie bekommen auf der nächsten Folie eine Halbtestreliabilität vorgegeben. Berechnen Sie diese zunächst händisch. Im Anschluss daran nutzen Sie den Code zur Übungsaufgabe 1 in `10-Rel.R` und überprüfen ihr Ergebnis.

- Zeit: 10 Minuten
- Replikation: `set.seed(123)`
- Tipp: `Rel_SBK(X_p, X_q)`
- Anmerkung: Konzepte verstehen $\gg$ Codes verstehen!

Übungsaufgabe 2: Material

Nach untenstehender Zufallsaufteilung der Items ist folgende Halbwertsreliabilität geben:

```
X <- data.frame(MASS::mvrnorm(N, mu, Sigma))
m <- length(X) ; seq <- seq(m)
rseq <- sample(seq, m, replace=FALSE)
X_p <- X[,rseq[1:4]]
X_q <- X[,rseq[5:8]]
rsx_p <- rowSums(X_p)
rsx_q <- rowSums(X_q)
(Rel_halb <- cor(rsx_p, rsx_q))
```

```
## [1] 0.8746126
```

Übungsaufgabe 2: Lösungsvorschlag (..händisch & R)

$$
\begin{aligned}
Rel(X_{voll.}) &= \frac{2 \cdot Corr(X_p, X_q)}{1 + Corr(X_p, X_q)} \\
&= \frac{2 \cdot Rel_{halb}}{1 + Rel_{halb}} \\
&\approx \frac{2 \cdot 0.86}{1 + 0.86} \\
&\approx \frac{2 \cdot 0.86}{1 + 0.86} \\
&\approx \frac{1.72}{1.86} \\
&\approx \frac{1.72}{1.86} \\
&\approx 0.92
\end{aligned}
$$

Übungsaufgabe 2: Lösungsvorschlag (..in R)

```
Rel_SBK(X_p, X_q)
```

```
## [1] 0.9331129
```

Test

- Zusätzlich: Wichtig, dass Sie SPF mit der Hand ausrechnen können.

1. Aufgabe: Ausrechnen
2. Mit R überprüfen
3. An der Tafel zeigen
4. Odd-Even: Ungerade Items in eine Hälfte, gerade Items in die andere Hälfte sinnvoll bei Leistungstests mit Items, die in ihrer Schwierigkeit ansteigen
5. Zeitpartitionierungsmethode: Items werden nach der Bearbeitungszeit aufgeteilt sinnvoll bei Tests mit vielen gleichartigen Items
6. Itemzwillinge: Bildung von Itempaaren anhand von Schwierigkeit und Trennschärfe, zufällige Zuweisung zu einer Testhälfte sinnvoll bei heterogenen Items
7. Zufällige Aufteilung

Voraussetzungen: Test-Test Reliabilitäten

Section 3

Einflussfaktoren auf die Höhe der Reliabilität

Section 4

Anwendung: Konfidenzintervalle

Section 5

Selbststudium

Selbststudium

Nachfolgend haben Sie die Möglichkeit Datenmatrizen aus einer multivariaten Normalverteilung zu simulieren. Dazu benötigen Sie allerdings das Package `MASS`. Wenn Sie es nicht selbst installieren wollen, kopieren Sie den Code auf der nächsten Folie. Der Code-Schnipsel klärt, ob Sie das Package bereits installiert haben und bietet Ihnen gegebenenfalls an, es zu installieren. Nach erfolgreicher Installation, testen Sie mit der gleichen Funktion, ob alles passt.

Preparation

```r
if(!requireNamespace("MASS", quietly = TRUE)) {
  msg <- "'MASS' is not installed, want to install it? Type 'y
  answer <- readline(prompt = message(msg))
  no_msg <- "Did not install the package `MASS`."
  switch(answer,
         yes = install.packages("MASS"),
         no = stop(no_msg, call. = FALSE),
         stop("Please answer 'yes' or 'no'." ))
} else {
 message("`MASS is already installed!") ; Sys.sleep(1)
 message("Time to rock!\n(*weird guitar sound*)")
}
```

```
## `MASS is already installed!

## Time to rock!
## (*weird guitar sound*)
```

Paralleles Messmodell