Testtheorie und Testkonstruktion

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(6) Interne Konsistenz

Motivation

BDI-II Interne Konsistenz (Hautzinger, Keller, and Kühner (2006))

In Tabelle 2 sind bisherige Ergebnisse zur internen Konsistenz (Cronbach's α) des BDI II aufgelistet. In psychiatrischen Stichproben lag die interne Konsistenz im Bereich von $0.89 \leq \alpha \leq 0.94$, in nicht klinischen Stichproben im Bereich von $0.84 \leq \alpha \leq 0.91$.

Tabelle 2: Ergebnisse zur internen Konsistenz des BDI II aus internationalen Studien

Studie	Stichprobe	Cronbach's α
Psychiatrische Stichproben		
Beck et al. (1996)	500 ambulante Patienten	.92
Beck, Steer et al. (1996)	140 ambulante Patienten	.91
Steer et al. (1998)	840 ambulante Patienten	.92
Steer et al. (1999)	210 ambulante Patienten	.90
Steer et al. (2000)	130 stationäre Geriatriepatienten	.89
Arnau et al. (2001)	340 ambulante Patienten	.94
Krefetz et al. (2002)	100 stationäre jugendliche Patienten (12-17 Jahre)	.92
Kumar et al. (2002)	100 stationäre jugendliche Patienten (12-17 Jahre)	.94
Krefetz et al. (2003)	200 jugendliche Patienten (13-17 Jahre)	.89
Nichtpsychiatrische Stichproben		
Osman et al. (1997)	230 Studenten	.90
Steer & Clark (1997)	160 Studenten	.89
Dozios et al. (1999)	1022 Studenten	.91
Whisman et al. (2000)	576 Studenten	.89
Leigh&Anthony-Tolbert (2001)	36 taube Studenten	.88
Grothe et al. (2005)	220 medizinische Ambulanzpatienten	.90
Nichtenglischsprachige Stichproben		
Al-Musawi (2001)	200 arabische Studenten	.84
Coelho et al. (2002)	775 portugiesische Erwachsene	.89
Kojima et al. (2002)	766 japanische Erwachsene	.87
Bonilla et al. (2004)	351 spanische Studenten	.88
Ghassemzadeh et al. (2005)	125 iranische Studenten	.87

Motivation

Deutsche Version des BDI-II Interne Konsistenz (Hautzinger, Keller, and Kühner (2006))

Die internen Konsistenzwerte (Cronbach's α) des BDI II sind, getrennt für die jeweiligen Teilstichproben, in Tabelle 7 aufgeführt. Die entsprechenden Werte liegen alle über 0.89 und damit im Bereich dessen, was in den internationelen Studien gefunden wurde (Tabelle 2). Die hier identifizierten hohen Konsistenzwerte des deutschen BDI II lassen auf eine hohe Homogenität des Verfahrens schließen , die es rechtfertigt, den Summenwert als Maß der Depressionsschwere zu verwenden.

Tabelle 7: Konsistenzschätzungen (Cronbach's alpha) für verschiedene Stichproben

Stichprobe	Cronbach's α
Depressive Patienten in Behandlung	.93
Entlassene, behandelte, gebesserte ehemals Depressive	.89
Patienten mit primär anderen psychischen Störungen	.92
Internetstichprobe	.91
Gesunde	.90

m-**Komponententestmodelle**

Cronbach's α

Selbstkontrollfragen

Vorbemerkungen

υ_{ij} bezeichnete bisher die Zufallsvariable zur Modellierung des Observed Scores der $j$ten Testmessung der $i$ten Person, wobei wir m Testmessungen $j = 1, ..., m$ und n Personen $i = 1, ..., n$ angenommen haben.

$\Rightarrow$ Wir haben bisher offen gelassen, ob mit der *$j$ten Testmessung* ein *Test* oder ein *Item* gemeint ist.

$\Rightarrow$ In diesem Abschnitt identifizieren wir die $j$te Testmessung nun mit dem $j$ten Item *eines* Tests.

υ_{ij} bezeichnet also nun die Zufallsvariable zur Modellierung des Observed Scores des $j$ten Items der $i$ten Person in einem Test, wobei wir weiterhin m Items $j = 1, ..., m$ und n Personen $i = 1, ..., n$ annehmen.

Insbesondere gehen wir davon aus, dass für jede Person ein Gesamt-Observed-Score durch Summation über die Items eines Test gebildet wird. Die Zufallsvariable zur Modellierung dieses Gesamt-Observed-Scores bezeichnen wir mit

$$\upsilon_i := \sum_{j=1}^{m} \upsilon_{ij} \tag{1}$$

Wir fassen diese Vorüberlegungen in der Definition des *m-Komponententestmodells* zusammen.

Definition (m-Komponententestmodell)

Gegeben sei das Modell multipler Testmessungen für $i = 1, ..., n$ Personen und $j = 1, ..., m$ Testmessungen. Für $i = 1, ..., n$ seien

- $\upsilon_i := \sum_{j=1}^m \upsilon_{ij}$ die *Observed-Score-Summe* und
- $\tau_i := \sum_{j=1}^m \tau_{ij}$ die *True-Score-Summe*.

Dann heißt die gemeinsame Verteilung der υ_i und τ_i für $i = 1, ...n$ mit der Faktorisierungseigenschaft

$$\mathbb{P}(\tau_1, ..., \tau_n, \upsilon_1, ..., \upsilon_n) := \prod_{i=1}^n \mathbb{P}(\tau_i, \upsilon_i) = \prod_{i=1}^n \mathbb{P}(\upsilon_i|\tau_i)\mathbb{P}(\tau_i) \quad (2)$$

das m-Komponententestmodell, wenn gilt dass

$$\mathbb{P}(\tau_1, \upsilon_1) = \cdots = \mathbb{P}(\tau_n, \upsilon_n). \quad (3)$$

Definition (Reliabilität von m-Komponententestmodellen)

Gegeben sei ein m-Komponententestmodell mit Observed-Score-Summe υ_i und True-Score-Summe τ_i für $i = 1, ..., n$ Personen oder ein m-Parallelkomponententestmodell mit Observed-Score-Summe υ_i und True-Score τ_i für $i = 1, ..., n$. Dann ist die Reliabilität des Modells definiert als

$$\mathrm{R} := \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(\upsilon_i)} \text{ für ein beliebiges } 1 \le i \le n. \tag{4}$$

Bemerkung

- Die Definition ist die erste Umformung der in der vorherigen Einheit genutzten Definition.

m-Komponententestmodelle

Cronbach's α

Selbstkontrollfragen

Definition (Cronbach's α)

Gegeben sei ein m-Komponententestmodell. Dann heißt

$$\alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}(\upsilon_{ij})}{\mathbb{V}(\upsilon_i)}\right) \tag{5}$$

Cronbach's α oder *Koeffizient* α.

Bemerkungen

- $\mathbb{V}(\upsilon_i)$ ist die Varianz der Observed-Score-Summe für Personen $i = 1, ..., n$.
- $\mathbb{V}(\upsilon_{ij})$ ist die Varianz des Observed-Scores für Personen $i = 1, ..., n$ und Items $j = 1, ..., m$

Theorem (Cronbach's α und Reliabilität)

Gegeben sei ein m-Komponententestmodell mit Reliabilität R. Dann gilt für Cronbach's α, dass

$$\alpha \leq \mathrm{R} \tag{6}$$

und Gleichheit tritt insbesondere dann ein, wenn die Testmessungen des m-Komponententestmodell parallel sind.

Bemerkungen

- Cronbach's α ist eine untere Grenze für die Reliabilität eines m-Komponententestmodells.
- Die Relibabilität eines m-Komponententestmodells ist mindestens so groß wie Cronbach's α.
- Die Reliabilität eines m-Komponententestmodells kann größer als Cronbach's α sein.
- Für parallele Testmessungen (Items) ist die Reliabilität eines m-Komponententestmodells gleich α.

Cronbach's α

Beweis

Zur Vereinfachung der Notation verzichten wir auf die explizite Auszeichung der Personen $i = 1, ..., n$ und setzen

$$\upsilon_j := \upsilon_{ij}, \tau_j := \tau_{ij}, \upsilon := \sum_{j=1}^{m} \upsilon_j \text{ und } \tau := \sum_{j=1}^{m} \tau_j, \tag{7}$$

sowie

$$\mathrm{R} := \frac{\mathbb{V}(\tau)}{\mathbb{V}(\upsilon)} \text{ und } \alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}(\upsilon_j)}{\mathbb{V}(\upsilon)}\right). \tag{8}$$

Wir gehen in vier Schritten vor.

(1) (*Summendarstellung*) Wir schreiben zunächst die Summe von m Zahlen um. Speziell gilt für reelle Zahlen $x_1, ..., x_m$, dass

$$\sum_{j=1}^{m} x_j = \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} (x_j + x_k). \tag{9}$$

Beweis

Anstelle eines Beweises betrachten wir den Fall $m := 4$. Dann gilt

$$
\begin{aligned}
\frac{1}{3}\sum_{j=1}^{3}\sum_{k=j+1}^{4}(x_j + x_k) &= \frac{1}{3}\left(\sum_{k=1+1}^{4}(x_1 + x_k) + \sum_{k=2+1}^{4}(x_2 + x_k) + \sum_{k=3+1}^{4}(x_3 + x_k)\right) \\
&= \frac{1}{3}\left(\sum_{k=2}^{4}(x_1 + x_k) + \sum_{k=3}^{4}(x_2 + x_k) + \sum_{k=4}^{4}(x_3 + x_k)\right) \\
&= \frac{1}{3}\left((x_1 + x_2) + (x_1 + x_3) + (x_1 + x_4) + (x_2 + x_3) + (x_2 + x_4) + (x_3 + x_4)\right) \\
&= \frac{1}{3}\left((x_1 + x_1 + x_1) + (x_2 + x_2 + x_2) + (x_3 + x_3 + x_3) + (x_4 + x_4 + x_4)\right) \\
&= \frac{1}{3}\left(3x_1 + 3x_2 + 3x_3 + 3x_4\right) \\
&= x_1 + x_2 + x_3 + x_4 \\
&= \sum_{j=1}^{4} x_j
\end{aligned}
\tag{10}
$$

Beweis (fortgeführt)

(2) (*True-Score-Kovarianzungleichung*) Wir leiten nun im Modell multipler Testmessungen eine Ungleichung her. Dazu betrachten wir im Modell multipler Testmessungen die Varianz der Differenz zweier True-Scores τ_j und τ_k. Mit der Nicht-Negativität der Varianz und dem Theorem zur Varianz spezieller Linearkombinationen von Zufallsvariable in (2) Theoretische Grundlagen ergibt sich

$$\mathbb{V}(\tau_j - \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) - 2\mathbb{C}(\tau_j, \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) \geq 2\mathbb{C}(\tau_j, \tau_k) \tag{11}$$

Weiterhin ergibt sich für beliebige $1 \leq j, k \leq m$ parallele Testmessungen, dass

$$\mathbb{V}(\tau_j - \tau_k) = \mathbb{V}(f_j(\tau_1) - f_k(\tau_1)) = \mathbb{V}(\tau_1 - \tau_1) = \mathbb{V}(0) = 0. \tag{12}$$

In diesem Fall ergibt sich in obiger Ungleichung und ihrer Anwendung im Folgenden also Gleichheit.

(3) (*Summen-True-Score-Varianzungleichung*) Wir betrachten nun die Varianz der True-Score Summe im m-Komponententestmodell. Mit dem Theorem zur Varianz einer Linearkombination von Zufallsvariablen aus Einheit (2) Theoretische Grundlagen, der Summendarstellung aus (1) und der True-Score-Kovarianzungleichung aus (2) ergibt sich zunächst

Cronbach's α

Beweis (fortgeführt)

$$
\begin{aligned}
\mathbb{V}(\tau) &= \mathbb{V}\left(\sum_{j=1}^{m} \tau_j\right) \\
&= \sum_{j=1}^{m} \mathbb{V}(\tau_j) + 2\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \mathbb{C}(\tau_j, \tau_k) \\
&= \frac{1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \Big(\mathbb{V}(\tau_j) + \mathbb{V}(\tau_k)\Big) + 2\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \mathbb{C}(\tau_j, \tau_k) \\
&\geq \frac{1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) + \sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \left(\frac{1}{m-1} + 1\right)\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \mathbb{C}(\tau_j, \tau_k) \\
&= \left(\frac{1}{m-1} + \frac{m-1}{m-1}\right)\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \frac{1+m-1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \frac{m}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k)
\end{aligned}
\tag{13}
$$

Beweis (fortgeführt)

Mit Aussage (5) des Theorems zur Lokalen Unkorreliertheit des Modells multipler Testemessungen und wiederum mit dem Theorem zur Varianz einer Linearkombination von Zufallsvariablen aus Einheit (2) Theoretische Grundlagen gilt dann

$$
\begin{aligned}
\mathbb{V}(\tau) &\geq \frac{m}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \frac{m}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\upsilon_j, \upsilon_k) \\
&= \frac{m}{m-1}\left(\mathbb{V}\left(\sum_{j=1}^{m}\upsilon_j\right) - \sum_{j=1}^{m}\mathbb{V}\left(\upsilon_j\right)\right) \\
&= \frac{m}{m-1}\left(\mathbb{V}(\upsilon) - \sum_{j=1}^{m}\mathbb{V}\left(\upsilon_j\right)\right)
\end{aligned}
\tag{14}
$$

Cronbach's α

Beweis (fortgeführt)

(4) (*Reliabilität*) Wir betrachten schließlich die Reliabilität im m-Komponententestmodell. Es ergibt sich

$$
\begin{aligned}
\mathsf{R} &= \frac{\mathbb{V}(\tau)}{\mathbb{V}(\upsilon)} \\
&\ge \frac{\frac{m}{m-1}\left(\mathbb{V}(\upsilon) - \sum_{j=1}^{m} \mathbb{V}\left(\upsilon_j\right)\right)}{\mathbb{V}(\upsilon)} \\
&= \frac{m}{m-1}\left(\frac{\mathbb{V}(\upsilon)}{\mathbb{V}(\upsilon)} - \frac{\sum_{j=1}^{m} \mathbb{V}\left(\upsilon_j\right)}{\mathbb{V}(\upsilon)}\right) \\
&= \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}\left(\upsilon_j\right)}{\mathbb{V}(\upsilon)}\right) \\
&=: \alpha.
\end{aligned}
\tag{15}
$$

$\square$

Cronbach's α

Bemerkungen

- $\mathbb{V}(\upsilon_i)$ ist die Varianz der Observed-Score Summen für Personen $i = 1, ..., n$.
- Es gilt $\mathbb{V}(\upsilon_1) = \cdots = \mathbb{V}(\upsilon_n)$.
- Ein Schätzer für $\mathbb{V}(\upsilon_i)$ ist die Stichprobenvarianz der Observed-Score-Summe

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i - \bar{\upsilon})^2 \text{ mit } \bar{\upsilon} := \frac{1}{n}\sum_{i=1}^{n}\upsilon_i. \tag{16}$$

- $\mathbb{V}(\upsilon_{ij})$ ist die Varianz des Observed-Scores für Personen $i = 1, ..., n$ und Items $j = 1, ..., m$.
- Es gilt $\mathbb{V}(\upsilon_{1j}) = \cdots = \mathbb{V}(\upsilon_{nj})$ für $i = 1, ..., n$.
- Ein Schätzer für $\mathbb{V}(\upsilon_{ij})$ ist die Stichprobenvarianz des Observed-Scores von Item j

$$S_j^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(\upsilon_{ij} - \bar{\upsilon}_j\right)^2 \text{ mit } \bar{\upsilon}_j := \frac{1}{n}\sum_{i=1}^{n}\upsilon_{ij}. \tag{17}$$

Cronbach's α

Simulation im Modell paralleler Testmessungen

$$n := 30, m := 21 \text{ mit } \mathbb{P}(\tau_i) := N(1,1), \mathbb{P}(v_{ij}|\tau_i) := N(\tau_i, 4) \text{ für } i = 1,...,n, j = 1,...,m. \tag{18}$$

```
library(psych)                                   # R Paket zur Testanalyse
set.seed(0)                                      # Reproduzierbarkeit
n   = 30                                         # Personenanzahl
m   = 21                                         # Itemanzahl
mu  = 1                                          # True-Score Erwartungswertparameter
T   = matrix(rep(NaN, n)   , nrow = n)           # True-Score Array
Y   = matrix(rep(NaN, n*m), nrow = n)            # Observed-Score Array
for(i in 1:n){                                   # Iteration über Personen
  T[i]  = rnorm(2,mu,sqrt(1))                    # True-Score Realisierung
  for(j in 1:m){                                 # Iteration über Items
    Y[i,j]  = rnorm(1,T[i],sqrt(4))}}            # Observed-Score Realisierung
vsi  = var(apply(Y,1,sum))                       # Stichprobenvarianz der Observed-Score-Summem
siv  = sum(apply(Y,2,var))                       # Summe der Item-Stichprobenvarianzen
a    = (m/(m-1))*(1-(siv/vsi))                   # direkte Berechnung von Cronbach's alphca
ap   = alpha(Y,warnings = F)                     # Berechnung von Cronbach's alpha mit psych
```

```
Cronbach's alpha (manuell) :  0.823
Cronbach's alpha (psych)   :  0.823
```

m-Komponententestmodelle

Cronbach's α

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition eines m-Komponententestmodells wieder.
2. Geben Sie die Definition der Reliabilität eines m-Komponententestmodells wieder.
3. Geben Sie die Definition von Cronbach's α wieder.
4. Geben Sie das Theorem zum Zusammenhang von Cronbach's α und der Reliabilität wieder.

Referenzen

Hautzinger, M, F. Keller, and C. Kühner. 2006. *BDI-II Beck Depressions-Inventar*. Pearson.