



Grundlagen der Testtheorie

BSc Psychologie WiSe 2025/26

Prof. Dr. Dirk Ostwald

(2) Klassische Testtheorie

Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

Motivation

Beispielaussagen zur Reliabilität des BDI-II

Retest-Reliabilität

In der Studie von Beck et al. (1996) resultierte bei 26 Patienten, die im Abstand von einer Woche untersucht wurden, ein **Retestkoeffizient** von $r_{tt} = .93$. Aus Studentenstichproben werden für ein- bis zweiwöchige Zeiträume Koeffizienten von $r_{tt} = .74 - .96$ berichtet (Leigh & Antony-Tolbert, 2001; Al-Musawi, 2001; Sprinkle, Lurie, Insko et al., 2002; Ghassemzadeh, Mojtabai, Karamghadiri & Ebrahimkhani, 2005).

Interne Konsistenz

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 $.89 \leq \alpha \leq .94$, in nichtklinischen Stichproben im Bereich von $.84 \leq \alpha \leq .91$.

Hautzinger et al. (2006)

Klassische Testtheorie

- Probabilistisches Modell von Item- und Summenwerten eines psychologischen Tests
- Zentrale Beiträge von Gulliksen (1950) und Lord & Novick (1968)
- Modell der Form

$$\text{Beobachteter Wert} = \text{Wahrer Wert} + \text{Messfehler} \quad (1)$$

- Häufig als “Messfehlertheorie” beschrieben
- Erklärung von Antwortmustern durch einen nicht-direkt observierbaren “wahren Wert”
- Im Gegensatz zum ALM wird auch der “wahre Wert” als Zufallsvariable modelliert
- Meist verteilungsfrei und mit einem Fokus auf Erwartungswerten und (Ko)Varianzen formuliert
- Modellierung von intra- und interindividueller Variabilität
- Spezialfall des faktoranalytischen Modells, enger Bezug zu Linear-Mixed-Effects-Modellen
- Grundlage der Schätzung von Reliabilität und interner Konsistenz psychologischer Tests

⇒ Retest-Reliabilität, Cronbach's α , Spearman-Brown-Formel

Standarddatenstruktur der Testtheorie und Fragebogendatenanalyse

Beobachtung von Werten von m Testmessungen von n Testpersonen

Table 1: 20 Testpersonen $\times$ 4 Testmessungen

Testperson	Testmessung 1	Testmessung 2	Testmessung 3	Testmessung 4
1	13	43	41	34
2	48	34	23	32
3	10	24	11	29
4	43	42	27	48
5	32	29	31	38
6	23	44	23	22
7	27	15	47	31
8	42	19	10	38
9	30	47	49	37
10	30	29	15	42
11	19	37	32	30
12	16	29	15	40
13	18	32	49	47
14	24	15	48	26
15	30	49	20	18
16	46	34	26	48
17	50	15	45	32
18	34	48	22	28
19	46	15	49	35
20	46	33	34	39

Definition (Wahrer Wert, beobachteter Wert und Messfehler)

Für $i = 1, \dots, n$ und $j = 1, \dots, m$ ist der *wahre Wert* (*true score*) t_{ij} einer Testperson i für eine Testmessung j definiert als der bedingte Erwartungswert des *beobachteten Werts* (*observed score*) der Testperson für diese Testmessung

$$t_{ij} := \mathbb{E}(y_{ij} | \tau_{ij} = t_{ij}). \quad (2)$$

Der *Messfehler* einer Testperson für diese Testmessung ist definiert als die Zufallsvariable

$$\varepsilon_{ij} := y_{ij} - t_{ij}. \quad (3)$$

Bemerkungen

- Ausgangsdefinition der klassischen Testtheorie nach Novick (1966) und Lord & Novick (1968).
- Eine *Testmessung* mag aus einem Item oder der Summe mehrerer Items bestehen.
- Erst im m -Komponentenmodell werden wir *Itemscores* und *Summenscores* unterscheiden.
- Man beachte, dass τ_{ij} , y_{ij} und ε_{ij} Zufallsvariablen sind
- Man beachte, dass $t_{ij} \in \mathbb{R}$ eine Konstante ist.
- Die Definition von t_{ij} ist zirkulär, da t_{ij} mithilfe von t_{ij} definiert wird.
- Die implizit vorhandene bedingte Verteilung $\mathbb{P}(y_{ij} | \tau_{ij} = t_{ij})$ wird *Propensitätsverteilung* genannt.
- Die Propensitätsverteilung modelliert intraindividuelle Variabilität beobachteter Werte bei festem wahren Wert.
- Die klassische Ausgangsdefinition findet sich weiterhin in vielen Lehrbüchern zur Testtheorie

Definition (Bedingter Erwartungswert)

Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2)$ mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$, WMF oder WDF $p(x_1, x_2)$ und bedingter WMF oder WDF $p(x_1|x_2)$ für alle $x_2 \in \mathcal{X}_2$. Dann ist der *bedingte Erwartungswert* von ξ_1 gegeben $\xi_2 = x_2$ definiert als

- $\mathbb{E}(\xi_1|\xi_2 = x_2) := \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2)$, wenn ξ ein diskreter Zufallsvektor ist.
- $\mathbb{E}(\xi_1|\xi_2 = x_2) := \int_{\mathcal{X}_1} x_1 p(x_1|x_2) dx_1$, wenn ξ ein kontinuierlicher Zufallsvektor ist.

Bemerkungen

- Man vergegenwärtige sich die Definitionen von bedingter Verteilung und Erwartungswert.
- Der bedingte Erwartungswert ist der Erwartungswert einer Zufallsvariable in einer bedingten Verteilung.
- Obige Definition definiert den bedingten Erwartungswert für einen festen Wert x_2 von ξ_2 .
- Bei einem festen Wert x_2 von ξ_2 ist $\mathbb{E}(\xi_1|\xi_2 = x_2)$ damit ein fester Wert.
- Allgemein ist $\mathbb{E}(\xi_1|\xi_2)$ allerdings eine Zufallsvariable, da ξ_2 eine Zufallsvariable ist.
- Analog definiert man bedingte Varianzen, bedingte Kovarianzen und bedingte Korrelationen.
- Bedingte Kovarianzen und bedingte Korrelationen waren im Kontext partieller Korrelationen Thema.

Bedingter Erwartungswert - Beispiel

Für einen zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$, der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ mit $\mathcal{X}_1 := \{1, 2, 3, 4\}$ und $\mathcal{X}_2 = \{1, 2, 3\}$ annahme, seien bedingte WMFen für $p(x_1|x_2)$ wie folgt gegeben:

$p(x_1 x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$
$p(x_1 = 1 x_2)$	$\frac{1}{4}$	$\frac{1}{3}$	0
$p(x_1 = 2 x_2)$	0	$\frac{2}{3}$	$\frac{1}{3}$
$p(x_1 = 3 x_2)$	$\frac{1}{2}$	0	$\frac{1}{3}$
$p(x_1 = 4 x_2)$	$\frac{1}{4}$	0	$\frac{1}{3}$

Dann ergeben sich die bedingten Erwartungswerte

$$\begin{aligned}\mathbb{E}(\xi_1|\xi_2 = 1) &= \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2 = 1) = 1 \cdot \frac{1}{4} + 2 \cdot 0 + 3 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{11}{4} \\ \mathbb{E}(\xi_1|\xi_2 = 2) &= \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2 = 2) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} + 3 \cdot 0 + 4 \cdot 0 = \frac{5}{3} \\ \mathbb{E}(\xi_1|\xi_2 = 3) &= \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2 = 3) = 1 \cdot 0 + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} = \frac{8}{3}.\end{aligned}\tag{4}$$

Motivation der klassischen Ausgangsdefinition nach Novick (1966) und Lord & Novick (1968)

- Novick (1966) und Lord & Novick (1968) wollten sich einer Diskussion zur Interpretation des wahren Werts entziehen. Hätten Sie den wahren Wert einer Testperson als Realisierung einer latenten Zufallsvariable definiert, so hätten sie sich der Frage stellen müssen, wie diese latente Variable zu interpretieren sei. Anstelle einer Diskussion zur metaphysischen Bedeutung der Zufallsvariable τ_{ij} haben sie eine operationalistische Definition des wahren Werts versucht, die auf beobachtbaren Größen basiert. Dabei ist die Kernidee, dass der wahre Wert einer Testperson für eine Testmessung der "Durchschnitt der beobachteten Werte" einer Testperson über "wiederholte Testmessungen unter identischen Bedingungen" ist.

Zur Bedeutung des wahren Werts zitieren Lord & Novick (1968), S. 29 - 30 Lazarsfeld (1959):

- "Angenommen, wir fragen eine Testperson, Herrn Brown, wiederholt, ob er die Vereinten Nationen befürwortet; nehmen wir weiter an, dass wir ihm nach jeder Frage "das Gehirn waschen" und ihm dann dieselbe Frage erneut stellen. Da Herr Brown unsicher ist, wie er zu den Vereinten Nationen steht, wird er manchmal eine befürwortende und manchmal eine ablehnende Antwort geben. Nachdem wir dieses Verfahren viele Male durchgeführt haben, berechnen wir anschließend den Anteil der Male, in denen Herr Brown die Vereinten Nationen befürwortet hat."

Diese Proportion wollen Novick (1966) und Lord & Novick (1968) dann als wahren Wert verstehen (vgl. auch Borsboom et al. (2004)).

Probleme der klassischen Ausgangsdefinition nach Novick (1966) und Lord & Novick (1968)

- Eine anti-metaphysische Grundhaltung ist nicht durchhaltbar: zum einen handelt es sich bei der “wiederholten Testmessung unter identischen Bedingungen” um ein idealisiertes, in der Realität nicht durchführbares, Gedankenexperiment. Zum anderen definieren Novick (1966) und Lord & Novick (1968) den wahren Wert auch gerade nicht als (endlichen) Mittelwert einer Messreihe, sondern als idealisierten Erwartungswert einer Zufallsvariable. Als operationalistische Definition überzeugt die klassische Ausgangsdefinition nicht.
- Die klassische Ausgangsdefinition die Entwicklung einer Standardmessfehlertheorie, wie sie beispielsweise die das allgemeine lineare Modell, für die Analyse von Tests und Fragebögen aber beträchtlich. Es werden keine Fehler angenommen, sondern diese ergeben sich als Differenz wahrer und beobachteter Werte. Diese, gegenüber den Standardtheorien “umgekehrt” aufgezugene Theorie mag einer der Gründe sein, warum die Klassische Testtheorie bis heute lediglich in der Psychologie und wenig darüber hinaus von Bedeutung ist.
- In der Benennung der impliziten Verteilung $\mathbb{P}(y_{ij}|\tau_{ij} = t_{ij})$ als *Propensitätsverteilung* klingt die Propensitätsinterpretation von Wahrscheinlichkeiten als kausal bedingte “Verwirklichungstendenzen”, die, im Gegensatz zur Frequentistischen Interpretation auch im Einzelfall Sinn ergeben, an. Allerdings unterstellen Propensitätsverteilungen kausale Prozesse, die in der Regel nicht spezifiziert und damit auch nicht beobachtbar sind, und führen damit letztlich auch wieder auf metaphysische Aussagen (vgl. auch Borsboom et al. (2004) und Borsboom (2009)).

Charakter der im Folgenden gewählten Formalisierung der Klassischen Testtheorie

- Ansatz des modell-basierten Realismus zur Modellierung eines $n \times m$ Datensatzes.
- Moderne Formulierung als Modell mit latenten (nicht-direkt beobachteten) Zufallsvariablen
- Explizite Nutzung von $t_{ij} := \mathbb{E}(y_{ij} | \tau_{ij} = t_{ij})$ in Beweisen.
- Ähnliche Darstellung z.B. in Foster (2020)

Überblick

- (1) Formulierung des *Modells multipler Testmessungen*
- (2) Formulierung eines *Vereinfachten Modells multipler Testmessungen*
- (3) Eigenschaften des Modells multipler Testmessungen *bezüglich einer Testmessung*
- (4) Eigenschaften des Modells multipler Testmessungen *bezüglich zweier Testmessungen*
- (5) Formulierung des *Modells paralleler Testmessungen*
- (6) *Reliabilität* von Testmessungen und parallelen Testmessungen.
- (7) Formulierung des *m-Komponententestmodells*
- (8) Cronbach's α und Gesamtreliabilität
- (9) Spearman-Brown-Formel

Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

Definition (Modell multipler Testmessungen)

Für $i = 1, \dots, n$ und $j = 1, \dots, m$ seien τ_{ij} eine Zufallsvariable, die den *wahren Wert* der i ten Testperson in der j ten Testmessung modelliere, und y_{ij} eine Zufallsvariable, die den *beobachteten Wert* der i ten Testperson in der j ten Testmessung modelliere. Dann nennen wir die gemeinsame Verteilung der τ_{ij} und y_{ij} mit der Faktorisierungseigenschaft

$$\mathbb{P}(\tau_{11}, y_{11}, \dots, \tau_{nm}, y_{nm}) := \prod_{i=1}^n \mathbb{P}(\tau_{i1}, \dots, \tau_{im}) \prod_{j=1}^m \mathbb{P}(y_{ij} | \tau_{ij}) \quad (5)$$

das *Modell multipler Testmessungen*, wenn gilt, dass

$$\mathbb{P}(\tau_{11}, \dots, \tau_{1m}) \prod_{j=1}^m \mathbb{P}(y_{1j} | \tau_{1j}) = \dots = \mathbb{P}(\tau_{n1}, \dots, \tau_{nm}) \prod_{j=1}^m \mathbb{P}(y_{nj} | \tau_{nj}). \quad (6)$$

Das Modell multipler Testmessungen

Bemerkungen

- Das Modell macht keine speziellen parametrischen Verteilungsannahmen.
- Die gemeinsame Verteilung der τ_{ij}, y_{ij} faktorisiert über Testpersonen $i = 1, \dots, n$.
- $\Leftrightarrow$ Unabhängigkeitsannahme über Testpersonen hinsichtlich wahrer und beobachteter Werte
- Für $i = 1, \dots, n$ faktorisiert die gemeinsame Verteilung der $\tau_{i1}, \dots, \tau_{im}$ über $j = 1, \dots, m$ nicht immer.
- $\Leftrightarrow$ Die wahren Werte einer Testperson können zusammenhängen (Parallelität, τ -Äquivalenz, ...)
- Für jedes $i = 1, \dots, n$ sind die y_{ij} für $j = 1, \dots, m$ bedingt unabhängig gegeben τ_{ij} .
- $\Leftrightarrow$ Der wahre Wert in Testmessung $k \neq j$ beeinflusst den beobachteten Wert in Testmessung j nicht.
- $\Leftrightarrow$ Der beobachtete Wert in Testmessung $k \neq j$ beeinflusst den beobachteten Wert in Testmessung j nicht.
- Die Marginalverteilungen $\mathbb{P}(\tau_{i1}, \dots, \tau_{im}) \prod_{j=1}^m \mathbb{P}(y_{ij}|\tau_{ij})$ sind identisch für $i = 1, \dots, n$.
- $\Leftrightarrow$ Die Daten einer Testperson werden als unabhängig realisiert aus einer "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1}, \dots, \tau_{\bullet m}) \prod_{j=1}^m \mathbb{P}(y_{\bullet j}|\tau_{\bullet j}). \quad (7)$$

- Das $\bullet$ Subskript soll hier die Unspezifität bezüglich einer Testperson symbolisieren.
- Für eine Testmessung, also für $m := 1$, gilt nach Definition des Modells multipler Testmessungen

$$\mathbb{P}(\tau_{11}, y_{11}, \dots, \tau_{n1}, y_{n1}) := \prod_{i=1}^n \mathbb{P}(\tau_{i1}, y_{i1}) = \prod_{i=1}^n \mathbb{P}(\tau_{i1}) \mathbb{P}(y_{i1}|\tau_{i1}). \quad (8)$$

- Dabei sind die gemeinsamen Marginalverteilungen $\mathbb{P}(\tau_{i1}, y_{i1})$ nach Definition identisch.
- Bei $m := 1$ entsprechen die wahren und beobachteten Werte also unabhängige Realisierungen aus

$$\mathbb{P}(\tau_{\bullet 1}) \mathbb{P}(y_{\bullet 1}|\tau_{\bullet 1}) \quad (9)$$

- Verzichtet man auf die Subskripte $\bullet 1$, so gelangt man zum *Vereinfachten Modell der Klassischen Testtheorie*.

Definition (Vereinfachtes Modell der Klassischen Testtheorie)

τ sei eine Zufallsvariable, die die Verteilung der *wahren Werte* zu einer Testmessung in einer Population beschreibe, und y sei eine Zufallsvariable, die die Verteilung der *beobachteten Werte* dieser Testmessung beschreibe. Dann heißt die gemeinsame Verteilung von τ und y ,

$$\mathbb{P}(\tau, y) = \mathbb{P}(\tau)\mathbb{P}(y|\tau) \quad (10)$$

das *vereinfachte Modell der Klassischen Testtheorie*.

Bemerkungen

- Gegenüber dem Modell multipler Testmessungen treten weniger Zufallsvariablen und Indizes auf.
- Die Redundanz der Gleichheit vieler Verteilungen im Modell multipler Testmessungen tritt nicht auf.
- n beobachtete Testwerte werden als Teilbeobachtungen von Ziehungen aus $\mathbb{P}(\tau, y)$ gedacht.
- Für die Unabhängigkeit über Testpersonen mag man auch schreiben

$$(\tau_1, y_1), \dots, (\tau_n, y_n) \sim \mathbb{P}(\tau, y). \quad (11)$$

- Die obige Notation ist analog zu beispielsweise der wohlbekannten Schreibweise

$$y_1, \dots, y_n \sim N(\mu, \sigma^2). \quad (12)$$

Das Modell multipler Testmessungen

- Ein Beispieldatensatz für $n = 5$ Testpersonen bei einer Testmessung ist etwa

i	τ_i	y_i
1	15	13
2	46	48
3	13	10
4	41	43
5	34	32

- Die τ_i Werte sind dabei "virtuell", also im Modell simulierbar, aber in der Anwendung nicht vorhanden.
- Viele wichtige Eigenschaften des Modells multipler Testmessungen ergeben sich schon basierend auf $\mathbb{P}(\tau, y)$.
- Das Modell sieht einfacher zu handhaben aus als das Modell multipler Testmessungen
- Man findet das Vereinfachte Modell der klassischen Testtheorie oft in Lehrbüchern.
- Aber: Wenn mehrere Personen und mehrere Testmessungen in Spiel kommen wird unklar, wovon die Rede ist.
- Aber: Schätzer der Modellparameter beruhen immer auf allen beobachteten Werten $y_{1j}, \dots, y_{nj}$.
- Wir werden nach Bedarf zwischen beiden Modellformulierungen hin und her wechseln.
- Die Modellformulierung in Sinne des Modells multipler Testmessungen ist unser Standardfall.

Theorem (Eigenschaften bezüglich einer Testmessung)

Gegeben sei das Modell multipler Testmessungen. Dann gelten für alle $i = 1, \dots, n$ und alle $j = 1, \dots, m$

$$(1) \mathbb{E}(\varepsilon_{ij} | \tau_{ij} = t_{ij}) = 0$$

$$(2) \mathbb{E}(\varepsilon_{ij}) = 0$$

$$(3) \mathbb{C}(\tau_{ij}, \varepsilon_{ij}) = 0$$

$$(4) \mathbb{V}(y_{ij}) = \mathbb{V}(\tau_{ij}) + \mathbb{V}(\varepsilon_{ij})$$

$$(5) \mathbb{C}(y_{ij}, \tau_{ij}) = \mathbb{V}(\tau_{ij})$$

Bemerkungen

- Die Aussagen beziehen sich auf die Verteilung einer Testperson i (und damit auf alle Personen $i = 1, \dots, n$).
- 1 bezieht sich auf die bedingte Verteilung von ε_{ij} bei einer Testmessung j .
- 2 - 5 beziehen sich auf Eigenschaften der Marginalverteilungen von $\tau_{ij}, y_{ij}, \varepsilon_{ij}$ bei einer Testmessung j .
- Im Sinne des Vereinfachten Modells der Klassischen Testtheorie werden obige Eigenschaften oft auch als

$$\mathbb{E}(\varepsilon | \tau = t) = 0, \quad \mathbb{E}(\varepsilon) = 0, \quad \mathbb{C}(\tau, \varepsilon) = 0, \quad \mathbb{V}(y) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) \text{ und } \mathbb{C}(y, \tau) = \mathbb{V}(\tau) \quad (13)$$

geschrieben.

Das Modell multipler Testmessungen

Bemerkung zu $\mathbb{E}(\varepsilon|\tau = t) = 0$ und $\mathbb{E}(\varepsilon) = 0$

- Die Tatsache, dass der bedingte Erwartungswert des Messfehlers 0 ist, folgt aus der Definition des wahren Werts. Dies ist ein anderer Ansatz als anzunehmen, dass der bedingte Erwartungswert des Messfehlers 0 ist. Weil der bedingte Erwartungswert des Messfehlers für jeden wahren Wert 0 ist, ist auch sein (marginaler) Erwartungswert 0.

Bemerkung zu $\mathbb{C}(\tau, \varepsilon) = 0$

- $\mathbb{C}(\tau, \varepsilon) = 0$ impliziert natürlich auch $\rho(\tau, \varepsilon) = 0$. Hohe oder niedrige wahre Werte sind also nicht systematisch mit hohen oder niedrigen Messfehlerwerten assoziiert.

Bemerkung zu $\mathbb{V}(y) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon)$

- Die Varianz beobachteter Werte kann additiv in Beiträge der Varianz des wahren Werts und des Messfehlers zerlegt werden. Diese Eigenschaft des Modells ist für eine Eigenschaft der Reliabilität essenziell.

Bemerkung zu $\mathbb{C}(y, \tau) = \mathbb{V}(\tau)$

- Die Kovarianz der beobachteten Werte und der wahren Werte einer Testmessung entspricht der Varianz der wahren Werte. Diese Eigenschaft des Modells ist für eine weitere Eigenschaft der Reliabilität essenziell.

Das Modell multipler Testmessungen

Beweis

Zum Beweis von (1) bis (5) setzen wir zur Vereinfachung der Notation im Sinne des Vereinfachten Modells der klassischen Testtheorie zunächst

$$y := y_{ij}, \tau := \tau_{ij} \text{ und analog } y := y_{ij}, t := t_{ij} \text{ sowie } \mathcal{Y} := \mathcal{Y}_{ij}, \mathcal{T} := \mathcal{T}_{ij} \quad (14)$$

für alle $i = 1, \dots, n$ und $j = 1, \dots, n$. Wir betrachten weiterhin (nur) den diskreten Fall, setzen also die Existenz einer Wahrscheinlichkeitsmassefunktion p der Form

$$p(t, y) = p(y|t)p(t) \quad (15)$$

voraus. Dabei schreiben wir Werte, die die Zufallsvariable τ annimmt als t und Werte, die die Zufallsvariable y annimmt der Einfachheit halber ebenfalls als y . Der kontinuierliche Fall folgt dann jeweils analog.

(1) Es gilt

$$\begin{aligned} \mathbb{E}(\varepsilon|\tau = t) &:= \mathbb{E}(y - \tau|\tau = t) \\ &= \sum_{y \in \mathcal{Y}} (y - t) p(y|t) \\ &= \sum_{y \in \mathcal{Y}} y p(y|t) - \sum_{y \in \mathcal{Y}} t p(y|t) \\ &= \mathbb{E}(y|\tau = t) - t \sum_{y \in \mathcal{Y}} p(y|t) \\ &= t - t \cdot 1 \\ &= 0. \end{aligned} \quad (16)$$

Beweis (fortgeführt)

(2) Es gilt

$$\begin{aligned}\mathbb{E}(\varepsilon) &:= \mathbb{E}(y - \tau) \\&= \sum_{y \in \mathcal{Y}} \sum_{t \in \mathcal{T}} (y - t) p(t, y) \\&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} (y - t) p(y|t) p(t) \\&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} y p(y|t) p(t) - t p(y|t) p(t) \\&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} p(t) (y p(y|t) - t p(y|t)) \\&= \sum_{t \in \mathcal{T}} p(t) \left(\sum_{y \in \mathcal{Y}} y p(y|t) - t \sum_{y \in \mathcal{Y}} p(y|t) \right) \\&= \sum_{t \in \mathcal{T}} p(t) (t - t \cdot 1) \\&= \sum_{t \in \mathcal{T}} p(t) \cdot 0 \\&= 0.\end{aligned}\tag{17}$$

Das Modell multipler Testmessungen

Beweis (fortgeführt)

(3) Es gilt

$$\begin{aligned}\mathbb{C}(\tau, \varepsilon) &= \mathbb{E}((\tau - \mathbb{E}(\tau))(\varepsilon - \mathbb{E}(\varepsilon))) \\&= \mathbb{E}((\tau - \mathbb{E}(\tau))\varepsilon) \\&= \mathbb{E}((\tau - \mathbb{E}(\tau))(y - \tau)) \\&= \sum_{y \in \mathcal{Y}} \sum_{t \in \mathcal{T}} ((t - \mathbb{E}(\tau))(y - t)) p(t, y) \\&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} (t - \mathbb{E}(\tau))(y - t) p(y|t) p(t) \\&= \sum_{t \in \mathcal{T}} p(t)(t - \mathbb{E}(\tau)) \sum_{y \in \mathcal{Y}} (y - t) p(y|t) \\&= \sum_{t \in \mathcal{T}} p(t)(t - \mathbb{E}(\tau)) \sum_{y \in \mathcal{Y}} (yp(y|t) - tp(y|t)) \\&= \sum_{t \in \mathcal{T}} p(t)(t - \mathbb{E}(\tau)) \left(\sum_{y \in \mathcal{Y}} yp(y|t) - t \sum_{y \in \mathcal{Y}} p(y|t) \right) \\&= \sum_{t \in \mathcal{T}} p(t)(t - \mathbb{E}(\tau)) (t - t \cdot 1) \\&= \sum_{t \in \mathcal{T}} p(t)(t - \mathbb{E}(\tau)) \cdot 0 \\&= 0.\end{aligned}\tag{18}$$

Das Modell multipler Testmessungen

Beweis (fortgeführt)

(4) Mit Theorem [Varianzen spezieller Linearkombinationen](#) sowie Aussage (3) des Theorems gilt

$$\mathbb{V}(y) = \mathbb{V}(\tau + \varepsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) + 2\mathbb{C}(\tau, \varepsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) + 2 \cdot 0 = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) \quad (19)$$

(5) Mit Theorem [Kovarianzverschiebungssatz](#), Theorem [Erwartungswert bei linear-affiner Kombination](#), Aussage (2) des Theorems und der Tatsache, dass mit Aussage (4) des Theorems außerdem folgt, dass

$$\mathbb{E}(\varepsilon\tau) = \mathbb{E}(\tau\varepsilon) = \mathbb{C}(\tau, \varepsilon) + \mathbb{E}(\tau)\mathbb{E}(\varepsilon) = 0 + \mathbb{E}(\tau) \cdot 0 = 0 \quad (20)$$

gilt

$$\begin{aligned} \mathbb{C}(y, \tau) &= \mathbb{E}(y\tau) - \mathbb{E}(y)\mathbb{E}(\tau) \\ &= \mathbb{E}((\tau + \varepsilon)\tau) - \mathbb{E}(\tau + \varepsilon)\mathbb{E}(\tau) \\ &= \mathbb{E}(\tau^2 + \varepsilon\tau) - (\mathbb{E}(\tau) + \mathbb{E}(\varepsilon))\mathbb{E}(\tau) \\ &= \mathbb{E}(\tau^2) + \mathbb{E}(\varepsilon\tau) - \mathbb{E}(\tau)^2 - \mathbb{E}(\varepsilon)\mathbb{E}(\tau) \\ &= \mathbb{E}(\tau^2) + \mathbb{E}(\varepsilon\tau) - \mathbb{E}(\tau)^2 - 0 \cdot \mathbb{E}(\tau) \\ &= \mathbb{E}(\tau^2) + 0 - \mathbb{E}(\tau)^2 - 0 \cdot \mathbb{E}(\tau) \\ &= \mathbb{E}(\tau^2) - \mathbb{E}(\tau)^2 \\ &= \mathbb{V}(\tau). \end{aligned} \quad (21)$$

□

Theorem (Normalverteilungsbeispiel)

Es sei $i = 1, \dots, n$, $m := 1$ mit

$$\mathbb{P}(\tau_{i1}) := N(\mu, 1) \text{ und } \mathbb{P}(y_{i1}|\tau_{i1}) := N(\tau_{i1}, 1) \quad (22)$$

Dann gelten

- (1) $\mathbb{P}(y_{i1}) = N(\mu, 2)$
- (2) $\mathbb{P}(\varepsilon_{i1}|\tau_{i1}) = N(0, 1)$
- (3) $\mathbb{P}(\varepsilon_{i1}) = N(0, 1)$
- (4) $\mathbb{C}(\tau_{i1}, \varepsilon_{i1}) = 0$

Bemerkungen

- Das Beispiel entspricht Exercise 2.17 in Lord & Novick (1968)
- Der Ansatz im Beispiel entspricht einer modernen Formulierung eines probabilistischen Modells.
- Die Beispieldefinitionen induzieren eine gemeinsame Normalverteilung von τ_{i1} und y_{i1} .
- Die Beispieldefinitionen und $\varepsilon_{i1} := y_{i1} - \mathbb{E}(y_{i1}|\tau_{i1})$ induzieren eine τ_{i1} bedingte Normalverteilung von ε_{i1} .
- Dies und die Beispieldefinitionen induzieren eine gemeinsame Normalverteilung von τ_{i1} und ε_{i1} .
- Die Aussagen folgen dann mit den [Eigenschaften gemeinsamer Normalverteilungen](#).

Das Modell multipler Testmessungen

Beweis

Zur Vereinfachung der Notation setzen wir $\tau := \tau_{i1}$, $y := y_{i1}$, $\varepsilon := \varepsilon_{i1}$.

(1) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(y|\tau) = N(\tau, 1) \quad (23)$$

induzierte gemeinsame Verteilung von τ und y , wobei offenbar

$$\mathbb{P}(y|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 1 \text{ und } b := 0 \quad (24)$$

gilt. Aus dem [Theorem zu gemeinsamen Normalverteilungen](#) ergibt sich dann zunächst, dass

$$\begin{pmatrix} \tau \\ y \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ 1 \cdot \mu + 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \cdot 1 \\ 1 \cdot 1 & 1 + 1 \cdot 1 \cdot 1 \end{pmatrix} \right) = N \left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right) \quad (25)$$

Aus dem [Theorem zu marginalen Normalverteilungen](#) ergibt sich dann durch Ablesen $y \sim N(\mu, 2)$.

(2) Wir betrachten $\mathbb{P}(\varepsilon|\tau = t)$ für einen beliebigen Wert $t \in \mathbb{R}$. Dann gilt

$$\varepsilon := y - t \text{ mit } y \sim N(t, 1) \quad (26)$$

Mit dem [Theorem zur linear-affinen Transformation normalverteilter Zufallsvektoren](#) gilt dann

$$\varepsilon \sim N(t - t, 1^2 \cdot 1) = N(0, 1) \quad (27)$$

Die Tatsache, dass dies für alle möglichen Werte von τ gilt, ist gerade Aussage (2).

Beweis (fortgesetzt)

(3) und (4) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(\varepsilon|\tau) = N(0, 1) \quad (28)$$

induzierte gemeinsame Verteilung von τ und ε , wobei offenbar

$$\mathbb{P}(\varepsilon|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 0 \text{ und } b := 0 \quad (29)$$

gilt. Aus dem [Theorem zu gemeinsamen Normalverteilungen](#) ergibt sich dann zunächst, dass

$$\begin{pmatrix} \tau \\ \varepsilon \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ 0 \cdot \mu + 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \cdot 0 \\ 0 \cdot 1 & 1 + 0 \cdot 1 \cdot 0 \end{pmatrix} \right) = N \left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \quad (30)$$

Aus dem [Theorem zu marginalen Normalverteilungen](#) ergeben sich dann durch Ablesen

$$\varepsilon \sim N(0, 1) \text{ und } \mathbb{C}(\tau, \varepsilon) = 0. \quad (31)$$

Das Modell multipler Testmessungen

Simulation des Normalverteilungsbeispiels mit $\mu := 1$ für $n := 10^4$

```
n      = 1e4                # Personenanzahl
m      = 1                  # Testmessungsanzahl
mu     = 1                  # Erwartungswertparameter wahrer Wert
T      = matrix(rep(NaN, n*m), nrow = n)  # Wahrer Wert Array
Y      = matrix(rep(NaN, n*m), nrow = n)  # Beobachteter Wert Array
E      = matrix(rep(NaN, n*m), nrow = n)  # Messfehler Array

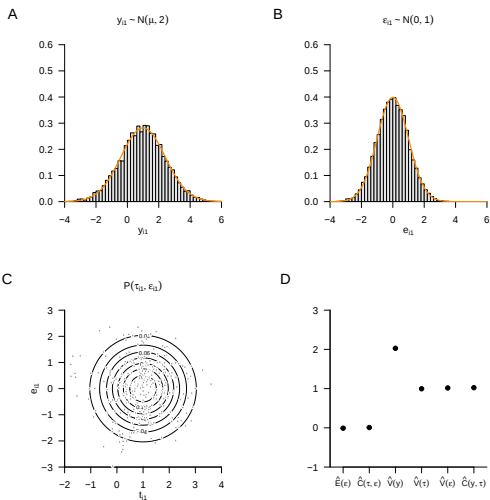
for(i in 1:n){              # Iteration über Testpersonen
  for(j in 1:m){             # Iteration über Testmessungen
    T[i,j] = rnorm(1,mu,1)    # Wahrer Wert Realisierung
    Y[i,j] = rnorm(1,T[i,j],1) # Beobachteter Wert Realisierung
    E[i,j] = Y[i,j] - T[i,j]} # Messfehler Realisierung
  }
e_hat_es = mean(E[,1])       # Erwartungswertschätzung Messfehler
c_hat_ts_es = cov(T[,1],E[,1]) # Kovarianzschätzung wahren Wert, Messfehler
v_hat_os = var(Y[,1])        # Varianzschätzung beobachteter Wert
v_hat_ts = var(T[,1])        # Varianzschätzung wahrer Wert
v_hat_es = var(E[,1])        # Varianzschätzung Messfehler
c_hat_os_ts = cov(Y[,1],T[,1]) # Kovarianzschätzung beobachteter Wert, wahrer Wert
```

Panels der untenstehenden Abbildung der Simulation

- A) Analytische und simulierte marginale Verteilung von y_{i1}
- B) Analytische und simulierte marginale Verteilung von ε_{i1}
- C) Analytische und simulierte gemeinsame Verteilung von τ_{i1} und ε_{i1} .
- D) Simulationsbasierte Schätzer relevanter Erwartungswerte, Varianzen und Kovarianzen

Das Modell multipler Testmessungen

Simulation des Normalverteilungsbeispiels mit $\mu := 1$ für $n := 10^4$



Das Modell multipler Testmessungen

Eine Linear Mixed Model Version des Normalverteilungsbeispiels

Wir betrachten das obige Normalverteilungsbeispiel konkret für $n := 5$ Testpersonen. Dann gilt für die gemeinsame WDF von $\tau_{11}, y_{11}, \dots, \tau_{51}, y_{51}$ nach Definition des Modells multipler Testmessungen

$$p(\tau_{11}, \tau_{21}, \tau_{31}, \tau_{41}, \tau_{51}, y_{11}, y_{21}, y_{31}, y_{41}, y_{51}) = \prod_{i=1}^5 p(\tau_{i1}) \prod_{i=1}^5 p(y_{i1}|\tau_{i1}) \quad (32)$$

Mit

$$\prod_{i=1}^5 p(\tau_{i1}) = N \left(\begin{pmatrix} \tau_{11} \\ \tau_{21} \\ \tau_{31} \\ \tau_{41} \\ \tau_{51} \end{pmatrix}; 1_5 \mu, I_5 \right) \Leftrightarrow \tau_{i1} = \mu + \eta_i \text{ mit } \eta_i \sim N(0, 1) \text{ u.i.v. für } i = 1, \dots, 5 \quad (33)$$

und

$$\prod_{i=1}^5 p(y_{i1}|\tau_{i1}) = N \left(\begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{51} \end{pmatrix}; \begin{pmatrix} \tau_{11} \\ \tau_{21} \\ \tau_{31} \\ \tau_{41} \\ \tau_{51} \end{pmatrix}, I_5 \right) \Leftrightarrow y_{i1} = \tau_{i1} + \varepsilon_{i1} \text{ mit } \varepsilon_{i1} \sim N(0, 1) \text{ u.i.v. für } i = 1, \dots, 5 \quad (34)$$

ergibt sich eine hierarchische Linear Mixed Model Darstellung.

Theorem (Eigenschaften bezüglich zweier Testmessungen)

Gegeben sei das Modell multipler Testmessungen. Dann gelten für alle Testpersonen $i = 1, \dots, n$ und alle Testmessungen j und k mit $1 \leq j, k \leq m$ und $j \neq k$, dass

$$(1) \quad \mathbb{C}(y_{ij}, y_{ik} | \tau_{ij} = t_{ij}, \tau_{ik} = t_{ik}) = 0,$$

$$(2) \quad \mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik} | \tau_{ij} = t_{ij}, \tau_{ik} = t_{ik}) = 0,$$

$$(3) \quad \mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik}) = 0,$$

$$(4) \quad \mathbb{C}(\tau_{ij}, \varepsilon_{ik}) = 0 \text{ und}$$

$$(5) \quad \mathbb{C}(y_{ij}, y_{ik}) = \mathbb{C}(\tau_{ij}, \tau_{ik}).$$

Bemerkungen

- Aussage (1) bezieht sich auf die bedingte gemeinsame Verteilung von y_{ij} und y_{ik} .
- Aussage (2) bezieht sich auf die bedingte gemeinsame Verteilung von ε_{ij} und ε_{ik} .
- Aussagen (3) bis (5) beziehen sich auf die jeweiligen Marginalverteilungen.
- Im Sinne des einfachen Modells der Klassischen Testtheorie werden obige Eigenschaften oft auch als

$$\mathbb{C}(y_j, y_k | \tau_j = t_j, \tau_k = t_k) = 0, \quad \mathbb{C}(\varepsilon_j, \varepsilon_k | \tau_j = t_j, \tau_k = t_k) = 0, \quad (35)$$

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = 0, \quad \mathbb{C}(\tau_j, \varepsilon_k) = 0, \quad \mathbb{C}(y_j, y_k) = \mathbb{C}(\tau_j, \tau_k) \quad (36)$$

geschrieben.

- Aussagen (1) - (4) werden manchmal unter dem Begriff der *Lokalen Unkorreliertheit* subsummiert.

Das Modell multipler Testmessungen

Bemerkung zu $\mathbb{C}(y_j, y_k | \tau_j = t_j, \tau_k = t_k) = 0$

- Gegeben die wahren Werte kovariieren die beobachteten Werte zweier Testmessungen nicht.
- In der auf die wahren Werte bedingten Verteilung zweier beobachteten Werte kovariieren diese nicht.

Bemerkung zu $\mathbb{C}(\varepsilon_j, \varepsilon_k | \tau_j = t_j, \tau_k = t_k) = 0$

- Gegeben die wahren Werte kovariieren die Messfehler zweier Testmessungen nicht.
- In der auf die wahren Werte bedingten Verteilung zweier Messfehler kovariieren diese nicht.

Bemerkung zu $\mathbb{C}(\varepsilon_j, \varepsilon_k) = 0$

- Über alle wahren Werte hinweg kovariieren die Messfehler zweier Testmessungen nicht.
- In der marginalen Verteilung zweier Messfehler kovariieren diese nicht.

Bemerkung zu $\mathbb{C}(\tau_j, \varepsilon_k) = 0$

- Die wahren Werte einer Testmessung j und der Messfehler einer Testmessung k kovariieren nicht.

Bemerkung zu $\mathbb{C}(y_j, y_k) = \mathbb{C}(\tau_j, \tau_k)$

- Die (marginale) Kovarianz zweier Testmessungen entspricht der (marginalen) Kovarianz der wahren Werte.

Beweis

Zur Vereinfachung der Notation verzichten wir in den Beweisen auf das i Subskript. Wir betrachten weiterhin nur den diskreten Fall und setzen die Existenz der marginalen Wahrscheinlichkeitsmassefunktion

$$p(t_j, y_j, t_k, y_k) = p(t_j, t_k)p(y_j|t_j)p(y_k|t_k) \quad (37)$$

und folglich auch der bedingten Wahrscheinlichkeitsmassefunktion

$$p(y_j, y_k|t_j, t_k) = \frac{p(t_j, y_j, t_k, y_k)}{p(t_j, t_k)} = \frac{p(t_j, t_k)p(y_j|t_j)p(y_k|t_k)}{p(t_j, t_k)} = p(y_j|t_j)p(y_k|t_k) \quad (38)$$

voraus. Dabei bezeichnen wir mit t_j, t_k Werte, die die Zufallsvariablen τ_j, τ_k annehmen können und der Einfachheit halber mit y_j, y_k Werte, die die Zufallsvariablen y_j, y_k annehmen können. Der kontinuierliche Fall folgt dann wieder analog.

Beweis (fortgeführt)

(1) Es gilt

$$\begin{aligned}\mathbb{C}(y_j, y_k | \tau_j = t_j, \tau_k = t_k) &= \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} (y_j - \mathbb{E}(y_j | \tau_j = t_j))(y_k - \mathbb{E}(y_k | \tau_k = t_k)) p(y_j | t_j) p(y_k | t_k) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) \sum_{y_k \in \mathcal{Y}_k} (y_k - t_k) p(y_k | t_k) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) \left(\sum_{y_k \in \mathcal{Y}_k} y_k p(y_k | t_k) - t_k \sum_{y_k \in \mathcal{Y}_k} p(y_k | t_k) \right) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) (t_k - t_k \cdot 1) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) \cdot 0 \\&= 0.\end{aligned}\tag{39}$$

Das Modell multipler Testmessungen

Beweis (fortgeführt)

(2) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j \varepsilon_k | \tau_j = t_j, \tau_k = t_k)$. Es gilt

$$\begin{aligned}\mathbb{E}(\varepsilon_j \varepsilon_k | \tau_j = t_j, \tau_k = t_k) &= \mathbb{E}((y_j - \tau_j)(y_k - \tau_k) | \tau_j = t_j, \tau_k = t_k) \\&= \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} (y_j - t_j)(y_k - t_k) p(y_j | t_j) p(y_k | t_k) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) \sum_{y_k \in \mathcal{Y}_k} (y_k - t_k) p(y_k | t_k) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) \left(\sum_{y_k \in \mathcal{Y}_k} y_k p(y_k | t_k) - t_k \sum_{y_k \in \mathcal{Y}_k} p(y_k | t_k) \right) \quad (40) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) (t_k - t_k \cdot 1) \\&= \sum_{y_j \in \mathcal{Y}_j} (y_j - t_j) p(y_j | t_j) \cdot 0 \\&= 0\end{aligned}$$

Mit Theorem [Verschiebungssatz der bedingten Kovarianz](#) und Aussage (1) des Theorem [Eigenschaften bezüglich einer Testmessung](#) folgt dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k | \tau_j = t_j, \tau_k = t_k) = \mathbb{E}(\varepsilon_j \varepsilon_k | \tau_j = t_j, \tau_k = t_k) - \mathbb{E}(\varepsilon_j | \tau_j = t_j) \mathbb{E}(\varepsilon_k | \tau_k = t_k) = 0 - 0 \cdot 0 = 0. \quad (41)$$

Das Modell multipler Testmessungen

Beweis (fortgeführt)

(3) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j \varepsilon_k)$. Mit dem Beweis von Aussage (2) ergibt sich

$$\begin{aligned}\mathbb{E}(\varepsilon_j \varepsilon_k) &= \mathbb{E}((y_j - \tau_j)(y_k - \tau_k)) \\&= \sum_{t_j \in \mathcal{T}_j} \sum_{t_k \in \mathcal{T}_k} \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} (y_j - t_j)(y_k - t_k) p(y_j | t_j) p(y_k | t_k) p(t_j, t_k) \\&= \sum_{t_j \in \mathcal{T}_j} \sum_{t_k \in \mathcal{T}_k} \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} (y_j - \mathbb{E}(y_j | \tau_j = t_j))(y_k - \mathbb{E}(y_k | \tau_k = t_k)) p(y_j | t_j) p(y_k | t_k) p(t_j, t_k) \\&= \sum_{t_j \in \mathcal{T}_j} \sum_{t_k \in \mathcal{T}_k} \mathbb{E}((y_j - \tau_j)(y_k - \tau_k) | \tau_j = t_j, \tau_k = t_k) p(t_j, t_k) \\&= \sum_{t_j \in \mathcal{T}_j} \sum_{t_k \in \mathcal{T}_k} \mathbb{E}(\varepsilon_j \varepsilon_k | \tau_j = t_j, \tau_k = t_k) p(t_j, t_k) \\&= \sum_{t_j \in \mathcal{T}_j} \sum_{t_k \in \mathcal{T}_k} 0 \cdot p(t_j, t_k) \\&= 0\end{aligned} \tag{42}$$

Mit Theorem [Kovarianzverschiebungssatz](#) und Aussage (2) von Theorem [Eigenschaften bezüglich einer Testmessung](#) ergibt sich dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = \mathbb{E}(\varepsilon_j \varepsilon_k) - \mathbb{E}(\varepsilon_j) \mathbb{E}(\varepsilon_k) = 0 - 0 \cdot 0 = 0. \tag{43}$$

Das Modell multipler Testmessungen

Beweis (fortgeführt)

(4) Mit dem Kovarianzverschiebungssatz Theorem [Kovarianzverschiebungssatz](#) und Aussage (2) von Theorem [Eigenschaften bezüglich einer Testmessung](#) gilt

$$\begin{aligned}\mathbb{C}(\tau_j, \varepsilon_k) &= \mathbb{E}(\tau_j \varepsilon_k) - \mathbb{E}(\tau_j) \mathbb{E}(\varepsilon_k) \\&= \mathbb{E}(\tau_j \varepsilon_k) - \mathbb{E}(\tau_j) \cdot 0 \\&= \mathbb{E}(\tau_j (y_k - \tau_k)) \\&= \sum_{t_j \in \mathcal{T}_j} \sum_{t_k \in \mathcal{T}_k} \sum_{y_k \in \mathcal{Y}_k} t_j (y_k - t_k) p(y_k | t_k) p(t_j, t_k) \\&= \sum_{t_j \in \mathcal{T}_j} t_j \sum_{t_k \in \mathcal{T}_k} p(t_j, t_k) \sum_{y_k \in \mathcal{Y}_k} (y_k - t_k) p(y_k | t_k) \\&= \sum_{t_j \in \mathcal{T}_j} t_j \sum_{t_k \in \mathcal{T}_k} p(t_j, t_k) \left(\sum_{y_k \in \mathcal{Y}_k} y_k p(y_k | t_k) - t_k \sum_{y_k \in \mathcal{Y}_k} p(y_k | t_k) \right) \\&= \sum_{t_j \in \mathcal{T}_j} t_j \sum_{t_k \in \mathcal{T}_k} p(t_j, t_k) (t_k - t_k \cdot 1) \\&= \sum_{t_j \in \mathcal{T}_j} t_j \sum_{t_k \in \mathcal{T}_k} p(t_j, t_k) \cdot 0 \\&= 0.\end{aligned}\tag{44}$$

Beweis (fortgeführt)

(5) Wir halten zunächst fest, dass mit Aussage (4) durch Vertauschen der Indizes und Theorem ?? auch

$$\mathbb{C}(\tau_k, \varepsilon_j) = \mathbb{C}(\varepsilon_j, \tau_k) = 0 \quad (45)$$

gilt. Mit Theorem Kovarianz bei paarweiser Addition und Aussage (3) gilt dann

$$\begin{aligned} \mathbb{C}(y_j, y_k) &= \mathbb{C}(\tau_j + \varepsilon_j, \tau_k + \varepsilon_k) \\ &= \mathbb{C}(\tau_j, \tau_k) + \mathbb{C}(\tau_j, \varepsilon_k) + \mathbb{C}(\varepsilon_j, \tau_k) + \mathbb{C}(\varepsilon_j, \varepsilon_k) \\ &= \mathbb{C}(\tau_j, \tau_k) + 0 + 0 + 0 \\ &= \mathbb{C}(\tau_j, \tau_k). \end{aligned} \quad (46)$$

□

Das Modell multipler Testmessungen

Normalverteilungsbeispiel

Wir betrachten den Fall zweier Testmessungen $j = 1, 2$ im Modell multipler Testmessungen.

Für $i = 1, \dots, n$ seien

$$\mathbb{P}(\tau_{i1}, \tau_{i2}) = \mathbb{P}(\tau_{i2}|\tau_{i1})\mathbb{P}(\tau_{i1}) \quad (47)$$

mit

$$\mathbb{P}(\tau_{i1}) := N(1, 1) \text{ und } \mathbb{P}(\tau_{i2}|\tau_{i1}) := N(\tau_{i1} + 1, 1) \quad (48)$$

Die Verteilung des wahren Werts von Testperson i in Testmessung $j = 2$ hängt also explizit von der Verteilung des wahren Werts von Testperson i in Testmessung $j = 1$ ab.

Weiterhin seien

$$\mathbb{P}(y_{i1}|\tau_{i1}) := N(\tau_{i1}, 1) \text{ und } \mathbb{P}(y_{i2}|\tau_{i2}) := N(\tau_{i2}, 2) \quad (49)$$

Die Propensitätsverteilungen von Testperson i in Testmessung $j = 1$ unterscheide sich also von der von Testperson i in Testmessung $j = 2$.

Das Modell multipler Testmessungen

Simulation des Normalverteilungsbeispiels für $n := 10^5$

```
n      = 1e5                # Personenanzahl
m      = 2                  # Testmessungsanzahl
mu     = 1                  # Wahrer Werte Erwartungswertparameter
T      = matrix(rep(NaN, n*m), nrow = n)  # Wahrer Wert Array
Y      = matrix(rep(NaN, n*m), nrow = n)  # Beobachteter Wert Array
E      = matrix(rep(NaN, n*m), nrow = n)  # Messfehler Array

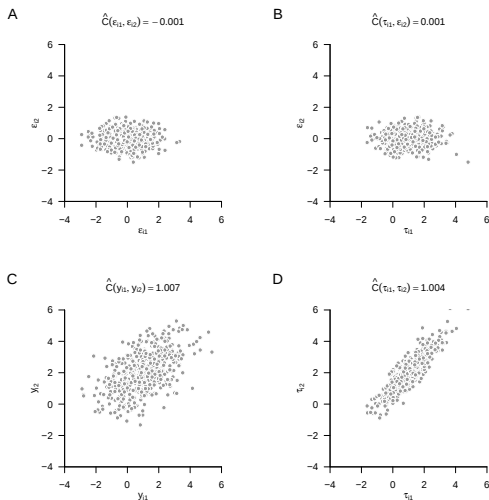
for(i in 1:n){              # Iteration über Testpersonen
  T[i,1] = rnorm(1,1,1)      # Wahrer Wert Realisierung      für j = 1
  Y[i,1] = rnorm(1,T[i,1],1) # Beobachteter Wert Realisierung für j = 1
  E[i,1] = Y[i,1] - T[i,1]  # Messfehler Realisierung      für j = 1
  T[i,2] = rnorm(1,T[i,1] + 1,.5) # Wahrer Wert Realisierung      für j = 2
  Y[i,2] = rnorm(1,T[i,2],.5) # Beobachteter Wert Realisierung für j = 2
  E[i,2] = Y[i,2] - T[i,2]}  # Messfehler Realisierung      für j = 2
c_hat_e1_e2 = cov(E[,1],E[,2]) # Kovarianzschätzung Messfehler 1, Messfehler 2
c_hat_t1_e2 = cov(T[,1],E[,2]) # Kovarianzschätzung wahrer Wert 1, Messfehler 2
c_hat_y1_y2 = cov(Y[,1],Y[,2]) # Kovarianzschätzung beobachteter Wert 1, beobachteter Wert 2
c_hat_t1_t2 = cov(T[,1],T[,2]) # Kovarianzschätzung wahrer Wert 1, wahrer Wert 2
```

Panels der untenstehenden Abbildung der Simulation

- A) Marginale Unkorreliertheit der Messfehler bezüglich zweier Testmessungen
- B) Marginale Unkorreliertheit des wahren Wertes und des Messfehlers bezüglich zweier Testmessungen
- C) Marginale Kovarianz der beobachteten Werte bezüglich zweier Testmessungen
- D) Marginale Kovarianz der wahren Werte bezüglich zweier Testmessungen

Das Modell multipler Testmessungen

Simulation des Normalverteilungsbeispiels für $n := 10^5$, 500 Stichprobenpunkte dargestellt



Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

Das Modell paralleler Testmessungen

Bisher haben wir im Modell der multiplen Testmessungen keine Aussage zu den Verhältnissen der True-Scores über verschiedene Testmessungen hinweg gemacht. Wir haben einerseits angenommen, dass für die Marginalverteilung der Testmessungen bei einer Person i keine Unabhängigkeit gelten muss, dass also im Allgemeinen gilt, dass für $i = 1, \dots, n$

$$\mathbb{P}(\tau_{i1}, \dots, \tau_{im}) \neq \prod_{j=1}^m \mathbb{P}(\tau_{ij}) \quad (50)$$

Andererseits haben wir die Form möglicher Abhängigkeiten zwischen den True-Scores $\tau_{i1}, \dots, \tau_{im}$ bislang nicht genauer spezifiziert.

Die Klassische Testtheorie betrachtet in dieser Hinsicht einige Spezialfälle, die sich im Allgemeinen durch funktionale Abhängigkeiten zwischen τ_{i1} und $\tau_{i2}, \dots, \tau_{im}$ der Form

$$\tau_{ij} = f(\tau_{i1}) \text{ für } f: \mathbb{R} \rightarrow \mathbb{R} \text{ und } j = 2, \dots, m. \quad (51)$$

auszeichnen. Wir betrachten im Folgenden den Fall, dass $f := \text{id}_{\mathbb{R}}$, dass also insbesondere für Realisierungen t_{ij} von τ_{ij} gilt, dass

$$t_{ij} = \text{id}_{\mathbb{R}}(t_{i1}) = t_{i1} \text{ für } j = 2, \dots, m, \quad (52)$$

dass also die Werte der True-Scores einer Person über Testmessungen identisch sind. Die Klassische Testtheorie bezeichnet solche Testmessungen als *parallele Testmessungen*. Eine weitere Form der funktionalen Abhängigkeit, die wir hier nicht weiter vertiefen wollen, ist der Fall, dass es sich bei f um eine linear-affine Funktion handelt, dass also

$$\tau_{ij} = f(\tau_{i1}) = a\tau_{i1} + b \text{ für } a, b \in \mathbb{R} \text{ und } j = 2, \dots, m. \quad (53)$$

Die Klassische Testtheorie bezeichnet solche Testmessungen als *wesentlich τ -äquivalente Testmessungen*.

Definition (Modell paralleler Testmessungen)

Für $i = 1, \dots, n$ und $j = 1, \dots, m$ sei

- τ_i eine Zufallsvariable, die den wahren Wert der i ten Testperson in jeder Testmessung $j = 1, \dots, m$ modelliere,
- y_{ij} eine Zufallsvariable, die den beobachteten Wert der i ten Testperson in der j ten Testmessung modelliere,
- $\varepsilon_{ij} := y_{ij} - \tau_i$ die Zufallsvariable, die den Messfehler der i ten Testperson in der j ten Testmessung modelliere.

Dann heißt die gemeinsame Verteilung der τ_i und y_{ij} mit den Faktorisierungseigenschaften

$$\mathbb{P}(\tau_1, y_{11}, \dots, y_{1m}, \dots, \tau_n, y_{n1}, \dots, y_{nm}) := \prod_{i=1}^n \mathbb{P}(\tau_i) \prod_{j=1}^m \mathbb{P}(y_{ij}|\tau_i) \quad (54)$$

das *Modell paralleler Testmessungen*, wenn gilt, dass

- (1) $\mathbb{P}(\tau_1) = \dots = \mathbb{P}(\tau_n)$,
- (2) $\mathbb{P}(y_{1j}|\tau_1) = \dots = \mathbb{P}(y_{nj}|\tau_n)$ für alle $1 \leq j \leq m$,
- (3) $\mathbb{E}(y_{ij}|\tau_i = t_i) = \mathbb{E}(y_{ik}|\tau_i = t_i) := t_i$ für alle $1 \leq i \leq n, 1 \leq j, k \leq m$,
- (4) $\mathbb{V}(y_{ij}|\tau_i = t_i) = \mathbb{V}(y_{ik}|\tau_i = t_i)$ für alle $1 \leq i \leq n, 1 \leq j, k \leq m$.

Bemerkungen

- Für Testperson i und Testmessungen $j = 1, \dots, m$ wird zunächst ein wahrer Wert t_i von $\mathbb{P}(\tau_i)$ realisiert.
- Für Testperson i und Testmessung j wird dann ein beobachteter Wert y_{ij} anhand von $\mathbb{P}(y_{ij} | \tau_i = t_i)$ realisiert.
- Die Werte des wahren Wertes einer Testperson werden über Testmessungen als identisch angenommen.
- Die Varianz der beobachteten Werte einer Testperson zwischen Messungen geht nur auf die Propensität zurück.
- Bei einer Testmessung entspricht das Modell paralleler Testmessungen dem Modell multipler Testmessungen.
- Das Theorem [Eigenschaften bezüglich einer Testmessung](#) gilt also analog.

Theorem (Eigenschaften des Modells paralleler Testmessungen)

Gegeben sei das Modell paralleler Testmessungen. Dann gelten für alle $i = 1, \dots, n$ und alle j, k mit $1 \leq j, k \leq m$ und $j \neq k$, dass

$$(1) \mathbb{E}(y_{ij}) = \mathbb{E}(y_{ik})$$

$$(2) \mathbb{V}(y_{ij}) = \mathbb{V}(y_{ik})$$

$$(3) \mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik}) = 0$$

$$(4) \mathbb{C}(\tau_i, \varepsilon_{ik}) = 0$$

$$(5) \mathbb{C}(y_{ij}, y_{ik}) = \mathbb{V}(\tau_i)$$

Bemerkungen

- Bei Paralleltestmessungen sind alle Erwartungswerte der beobachteten Werte einer Person identisch sind.
- Bei Paralleltestmessungen sind alle Varianzen der beobachteten Werte einer Person identisch sind.
- Aussagen (3) und (4) gelten analog im Modell multipler Testmessungen bezüglich zweier Testmessungen.
- Aussage (5) besagt insbesondere, dass $\mathbb{C}(y_{ij}, y_{ik})$ für beliebige j und k identisch zu $\mathbb{V}(\tau_i)$ sind.

Das Modell paralleler Testmessungen

Beweis

Zum Beweis setzen zur Vereinfachung der Notation zunächst

$$y_j := y_{ij}, y_k := y_{ik}, \tau := \tau_i, y_j := y_{ij}, y_k := y_{ik}, t := t_i, \mathcal{Y}_j := \mathcal{Y}_{ij}, \mathcal{Y}_k := \mathcal{Y}_{ik} \text{ und } \mathcal{T} := \mathcal{T}_i \quad (55)$$

für alle $i = 1, \dots, n$. Wir betrachten weiterhin (nur) den diskreten Fall, setzen also die Existenz einer Wahrscheinlichkeitsmassefunktion $p_{\tau, y}$ der Form

$$p(t, y_j, y_k) = p(y_j|t)p(y_k|t)p(t) \quad (56)$$

voraus und bezeichnen dabei mit t einen Wert, den die Zufallsvariable τ annimmt und der Einfachheit halber mit y_j, y_k Werte, die die Zufallsvariablen y_j, y_k annehmen können. Der kontinuierliche Fall folgt dann analog.

(1) Mit der Gleichheit der bedingten Erwartungswerte im Falle paralleler Testmessungen gilt

$$\mathbb{E}(y_j) = \sum_{t \in \mathcal{T}} \sum_{y_j \in \mathcal{Y}_j} y_j p(y_j|t)p(t) = \sum_{t \in \mathcal{T}} \mathbb{E}(y_j|\tau = t)p(t) = \sum_{t \in \mathcal{T}} \sum_{y_k \in \mathcal{Y}_k} y_k p(y_k|t)p(t) = \mathbb{E}(y_k).$$

(2) Mit Theorem **Iterierte Varianz** ergibt sich

$$\mathbb{V}(y_j) = \mathbb{V}(\mathbb{E}(y_j|\tau)) + \mathbb{E}(\mathbb{V}(y_j|\tau)) = \mathbb{V}(\mathbb{E}(y_k|\tau)) + \mathbb{E}(\mathbb{V}(y_k|\tau)) = \mathbb{V}(y_k).$$

Beweis (fortgeführt)

(3) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j \varepsilon_k)$. Mit Aussage (2) des Theorems zu den Eigenschaften des Modells multipler Testmessungen bei einer Testmessung gilt

$$\begin{aligned}\mathbb{E}(\varepsilon_j \varepsilon_k) &:= \mathbb{E}((y_j - \tau)(y_k - \tau)) \\&= \sum_{t \in \mathcal{T}} \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} (y_j - t)(y_k - t) p(t) p(y_j|t) p(y_k|t) \\&= \sum_{t \in \mathcal{T}} p(t) \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} (y_j - t)(y_k - t) p(y_j|t) p(y_k|t) \\&= \sum_{t \in \mathcal{T}} p(t) \sum_{y_j \in \mathcal{Y}_j} (y_j - t) p(y_j|t) \sum_{y_k \in \mathcal{Y}_k} (y_k - t) p(y_k|t) \\&= \sum_{t \in \mathcal{T}} p(t) \left(\sum_{y_j \in \mathcal{Y}_j} y_j p(y_j|t) - t \sum_{y_j \in \mathcal{Y}_j} p(y_j|t) \right) \left(\sum_{y_k \in \mathcal{Y}_k} y_k p(y_k|t) - t \sum_{y_k \in \mathcal{Y}_k} p(y_k|t) \right) \\&= \sum_{t \in \mathcal{T}} p(t) (t - t) (t - t) \\&= \sum_{t \in \mathcal{T}} p(t) \cdot 0 \cdot 0 \\&= 0.\end{aligned} \tag{57}$$

Das Modell paralleler Testmessungen

Beweis (fortgeführt)

Mit Theorem [Kovarianzverschiebungssatz](#) und wiederum mit Aussage (2) von Theorem [Eigenschaften bezüglich einer Testmessung](#) folgt dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = \mathbb{E}(\varepsilon_j \varepsilon_k) - \mathbb{E}(\varepsilon_j) \mathbb{E}(\varepsilon_k) = 0 - 0 \cdot 0 = 0 \quad (58)$$

(4) Mit Theorem [Kovarianzverschiebungssatz](#) und wiederum mit Aussage (2) von Theorem [Eigenschaften bezüglich einer Testmessung](#) folgt dann

$$\begin{aligned} \mathbb{C}(\tau, \varepsilon_k) &= \mathbb{E}(\tau \varepsilon_k) - \mathbb{E}(\tau) \mathbb{E}(\varepsilon_k) \\ &= \mathbb{E}(\tau \varepsilon_k) - \mathbb{E}(\tau) \cdot 0 \\ &= \mathbb{E}(\tau(y_k - \tau)) \\ &= \sum_{t \in \mathcal{T}} \sum_{y_k \in \mathcal{Y}_k} t(y_k - t) p(t) p(y_k | t) \\ &= \sum_{t \in \mathcal{T}} t p(t) \sum_{y_k \in \mathcal{Y}_k} (y_k - t) p(y_k | t) \\ &= \sum_{t \in \mathcal{T}} t p(t) \left(\sum_{y_k \in \mathcal{Y}_k} y_k p(y_k | t) - t \sum_{y_k \in \mathcal{Y}_k} p(y_k | t) \right) \\ &= \sum_{t \in \mathcal{T}} t p(t) (t - t) \\ &= \sum_{t \in \mathcal{T}} t p(t) \cdot 0 \\ &= 0. \end{aligned} \quad (59)$$

Beweis (fortgeführt)

(5) Wir halten zunächst fest, dass mit Aussage (2) des Theorems neben $\mathbb{C}(\tau, \varepsilon_k) = 0$ durch Austausch des Index und Theorem [Symmetrie der Kovarianz](#) auch

$$\mathbb{C}(\tau, \varepsilon_j) = \mathbb{C}(\varepsilon_j, \tau) = 0 \quad (60)$$

gilt. Mit Theorem [Kovarianz bei paarweiser Addition](#) und Aussage (1) des Theorems gilt dann

$$\begin{aligned} \mathbb{C}(y_j, y_k) &= \mathbb{C}(\tau + \varepsilon_j, \tau + \varepsilon_k) \\ &= \mathbb{C}(\tau, \tau) + \mathbb{C}(\tau, \varepsilon_k) + \mathbb{C}(\varepsilon_j, \tau) + \mathbb{C}(\varepsilon_j, \varepsilon_k) \\ &= \mathbb{C}(\tau, \tau) + 0 + 0 + 0 \\ &= \mathbb{C}(\tau, \tau) \\ &= \mathbb{V}(\tau). \end{aligned} \quad (61)$$

Das Modell paralleler Testmessungen

Normalverteilungsbeispiel

Wir betrachten den Fall zweier Testmessungen $j = 1, 2$ im Modell paralleler Testmessungen. Für $i = 1, \dots, n$ seien

$$\mathbb{P}(\tau_i) = N(1, 1) \text{ und } \mathbb{P}(y_{i1}|\tau_i) := \mathbb{P}(y_{i2}|\tau_i) := N(\tau_i, 1) \quad (62)$$

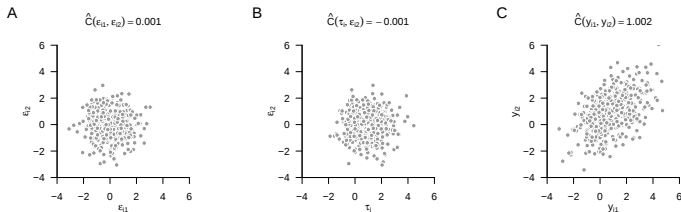
Für Testperson i gibt es also nur eine Zufallsvariable des wahren Werts für alle Testmessungen und die Propensitätsverteilungen unterscheiden sich zwischen Testmessungen nicht.

Simulation des Normalverteilungsbeispiels für $n = 10^5$

```
n      = 1e5                # Personenanzahl
m      = 2                  # Testmessungsanzahl
T      = matrix(rep(NaN, n) , nrow = n)  # Wahrer Wert Array
Y      = matrix(rep(NaN, n*m), nrow = n)  # Beobachteter Wert Array
E      = matrix(rep(NaN, n*m), nrow = n)  # Messfehler Array
for(i in 1:n){              # Personeniterationen
  T[i] = rnorm(1,1,1)        # Wahrer Wert Realisierung für j = 1,2
  for(j in 1:m){            # Testmessungsiterationen
    Y[i,j] = rnorm(1,T[i],1) # Beobachteter Wert Realisierung f
    E[i,j] = Y[i,j] - T[i]}  # Messfehler Realisierung
e_hat_o1_o2 = apply(Y, 2, mean) # Erwartungswertschätzung Beobachteter Wert 1, Beobachteter Wert 2
v_hat_o1_o2 = apply(Y, 2, var)  # Varianzzschätzung Beobachteter Wert 1, Beobachteter Wert 2
c_hat_e1_e2 = cov(E[,1],E[,2]) # Kovarianzzschätzung Messfehler 1, Messfehler 2
c_hat_t_e2 = cov(T      ,E[,2]) # Kovarianzzschätzung Wahrer Wert 1, Messfehler 2
c_hat_y1_y2 = cov(Y[,1],Y[,2])  # Kovarianzzschätzung Beobachteter Wert 1, Beobachteter Wert 2
v_hat_t     = var(T)            # Varianzzschätzung Wahrer Wert
```

Das Modell paralleler Testmessungen

Simulation des Normalverteilungsbeispiels für $n = 10^5$, 500 Stichprobenpunkte dargestellt



- A) Unkorreliertheit der Messfehler bezüglich zweier paralleler Testmessungen
- B) Unkorreliertheit des wahren Wertes und des Messfehlers bezüglich zweier paralleler Testmessungen
- C) Kovarianz der beobachteten Werte bezüglich zweier paralleler Testmessungen

Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

Testgütekriterium Reliabilität

“Ein Test ist dann *reliabel*, wenn er das Merkmal, das er misst, exakt, d.h. ohne Messfehler misst. Vor dem Hintergrund der Klassischen Testtheorie unterscheidet man die *Retest-Reliabilität*, die *Paralleltest-Reliabilität*, die *Testhalbierungs-Reliabilität* (*Split-Half-Reliabilität*) und die *Innere Konsistenz* eines Tests.”

Moosbrugger & Kelava (2012), Kapitel 2

“Unter *Reliabilität* wird die Genauigkeit einer Messung verstanden. Ein Testverfahren ist perfekt reliabel, wenn die damit erhaltenen Testwerte frei von zufälligen Messfehlern sind. Das Testverfahren ist umso weniger reliabel, je größer die Einflüsse von zufälligen Messfehlern sind”

Moosbrugger & Kelava (2012), Kapitel 6

Testgütekriterium Reliabilität

- Die Reliabilität ist definiert als die quadrierte Korrelation von beobachtetem und wahren Wert.
- Anhand der Eigenschaften des Modells multipler Testmessungen ergeben sich verschiedene Interpretationen.
- Die Interpretation nach Moosbrugger & Kelava (2012) finden sich auch.
- Messfehlerfreie Testmessungen sind allerdings sicherlich nur ein nicht zu erreichendes Ideal.
- Wir betrachten zunächst die Reliabilität von Testmessungen im Modell multipler Testmessungen.
- Allerdings ergibt sich dabei keine Möglichkeit, die Reliabilität von Testmessungen empirisch zu messen.
- Zentral ist dann die Definition der *Paralleltestreliabilität*. Diese ist empirisch messbar.
- Bei der Retest-Reliabilität handelt es sich um eine spezielle Paralleltestreliabilität.
- Innere Konsistenz und Split-Half-Reliabilität betrachten wir im Abschnitt Interne Konsistenz.

Definition (Reliabilität einer Testmessung)

Gegeben sei das Modell multipler Testmessungen für eine beliebige Testmessung j mit $1 \leq j \leq m$,

$$\mathbb{P}(\tau_{1j}, y_{1j}, \dots, \tau_{nj}, y_{nj}) := \prod_{i=1}^n \mathbb{P}(\tau_{ij}, y_{ij}) := \prod_{i=1}^n \mathbb{P}(y_{ij} | \tau_{ij}) \mathbb{P}(\tau_{ij}) \quad (63)$$

wobei nach Definition des Modells gilt, dass

$$\mathbb{P}(\tau_{1j}, y_{1j}) = \dots = \mathbb{P}(\tau_{nj}, y_{nj}). \quad (64)$$

Die *Reliabilität der Testmessung j* ist dann definiert als

$$R_j := \rho(y_{ij}, \tau_{ij})^2 \text{ für ein beliebiges } 1 \leq i \leq n. \quad (65)$$

Bemerkungen

- Vor dem Hintergrund des vereinfachten Modells der Klassischen Testtheorie schreibt man auch $R = \rho(y, \tau)^2$

Bemerkungen (fortgesetzt)

- Die Reliabilität einer Testmessung ist als quadrierte Korrelation von beobachtetem und wahren Wert definiert.
- Mit $-1 \leq \rho(y_{ij}, \tau_{ij}) \leq 1$ folgt direkt, dass $0 \leq R_j \leq 1$.
- $R_j = 0$ impliziert $\rho(y_{ij}, \tau_{ij}) = 0$, also die linear-affine Unabhängigkeit von beobachtetem und wahren Wert.
- $R_j = 0$ impliziert also, dass der beobachtete Wert hinsichtlich des wahren Werts nicht aussagekräftig ist
- $R_j = 1$ impliziert $\rho(y_{ij}, \tau_{ij}) = \pm 1$, also die linear-affine Abhängigkeit von beobachtetem und wahren Wert.
- $R_j = 1$ impliziert also, dass der beobachtete Wert hinsichtlich des wahren Werts aussagekräftig ist.
- Weil wahre Werte nur indirekt beobachtbar sind, wird eine alternative Form zu *Reliabilitätsschätzung* genutzt.

Theorem (Eigenschaften der Reliabilität einer Testmessung)

Gegeben sei das Modell multipler Testmessungen für eine beliebige Testmessung j mit $1 \leq j \leq m$,

$$\mathbb{P}(\tau_{1j}, y_{1j}, \dots, \tau_{nj}, y_{nj}) := \prod_{i=1}^n \mathbb{P}(y_{ij} | \tau_{ij}) \mathbb{P}(\tau_{ij}). \quad (66)$$

Dann gelten für die Reliabilität R_j der Testmessung j , dass

$$(1) \quad R_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})}$$

$$(2) \quad R_j = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}$$

für alle $1 \leq i \leq n$.

Beweis

(1) Mit Aussage (5) von Theorem [Eigenschaften bezüglich einer Testmessung](#) gilt

$$R_j = \rho(y_{ij}, \tau_{ij})^2 = \left(\frac{\mathbb{C}(y_{ij}, \tau_{ij})}{\mathbb{S}(y_{ij})\mathbb{S}(\tau_{ij})} \right)^2 = \frac{\mathbb{C}(y_{ij}, \tau_{ij})^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_{ij})} = \frac{\mathbb{V}(\tau_{ij})^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_{ij})} = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})}. \quad (67)$$

(2) Mit Aussage (4) von Theorem [Eigenschaften bezüglich einer Testmessung](#) gilt dann weiter

$$R_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij}) - \mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij})}{\mathbb{V}(y_{ij})} - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}. \quad (68)$$

□

Bemerkungen

- Da τ_{ij} und ε_{ij} nicht direkt beobachtbar sind, sind (1) und (2) nur von theoretischem Interesse.
- Die Reliabilität einer Testmessung ist der Anteil der Varianz des wahren Werts an der des beobachteten Werts,

$$R_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(\tau_{ij}) + \mathbb{V}(\varepsilon_{ij})}. \quad (69)$$

- Ist $\mathbb{V}(\tau_{ij}) = 0$, so ist auch $R_j = 0$, ist $\mathbb{V}(\varepsilon_{ij}) = 0$ so ist $R_j = 1$.
- $R_j > 0$ impliziert also, dass der wahre Wert eine von Null verschiedene Varianz (über Testpersonen) hat.
- Die Reliabilität entspricht auch 1 minus dem Anteil der Varianz des Messfehlers an der des beobachteten Werts.

Definition (Reliabilität einer Paralleltestmessung)

Gegeben sei das Modell paralleler Testmessungen für eine beliebige Testmessung j mit $1 \leq j \leq m$,

$$\mathbb{P}(\tau_1, y_{1j}, \dots, \tau_n, y_{nj}) := \prod_{i=1}^n \mathbb{P}(\tau_i, y_{ij}) = \prod_{i=1}^n \mathbb{P}(\tau_i) \mathbb{P}(y_{ij} | \tau_i) \quad (70)$$

wobei nach Definition des Modells gilt, dass

$$\mathbb{P}(\tau_1, y_{1j}) = \dots = \mathbb{P}(\tau_n, y_{nj}). \quad (71)$$

Die *Reliabilität der Paralleltestmessung j* ist dann definiert als

$$R_j := \rho(y_{ij}, \tau_i)^2 \text{ für ein beliebiges } 1 \leq i \leq n. \quad (72)$$

Bemerkungen

- In Bezug auf das vereinfachte Modell der Klassischen Testtheorie schreibt man auch hier $R = \rho(y, \tau)^2$

Theorem (Paralleltestreliabilität)

Gegeben sei das Modell paralleler Testmessungen. Dann gelten

- (1) $R_j = \frac{V(\tau_i)}{V(y_{ij})}$ für alle $1 \leq j \leq m$.
- (2) $R_j = 1 - \frac{V(\varepsilon_{ij})}{V(y_{ij})}$ für alle $1 \leq j \leq m$.
- (3) $R_j = \rho(y_{ij}, y_{ik}) = R_k$ für alle $1 \leq j, k \leq m$ mit $j \neq k$.

Bemerkungen

- Aussagen (1) und (2) sind analog zum Theorem der Reliabilität multipler Testmessungen.
- Aussage (3) des Theorems begründet die Paralleltest- und Retestreliabilitätsschätzung.
- Man sagt dazu vereinfachend, dass "die Korrelation paralleler Testmessungen gleich ihrer Reliabilität ist".
- Zur Messung der Reliabilität nutzt man eine Schätzung der Korrelation zweier paralleler Testmessungen.
- Die Klassische Testtheorie begründet dieses Vorgehen vor der Annahme von wahren und Messfehlerwerten.

Beweis

(1) Mit Aussage (5) Theorem [Eigenschaften des Modells paralleler Testmessungen](#) gilt

$$R_j = \rho(y_{ij}, \tau_i)^2 = \left(\frac{C(y_{ij}, \tau_i)}{S(y_{ij})S(\tau_i)} \right)^2 = \frac{C(y_{ij}, \tau_i)^2}{V(y_{ij})V(\tau_i)} = \frac{V(\tau_i)^2}{V(y_{ij})V(\tau_i)} = \frac{V(\tau_i)}{V(y_{ij})}. \quad (73)$$

(2) Mit Aussage (4) von Theorem [Eigenschaften bezüglich einer Testmessung](#) gilt

$$R_j = \frac{V(\tau_i)}{V(y_{ij})} = \frac{V(y_{ij}) - V(\varepsilon_{ij})}{V(y_{ij})} = \frac{V(y_{ij})}{V(y_{ij})} - \frac{V(\varepsilon_{ij})}{V(y_{ij})} = 1 - \frac{V(\varepsilon_{ij})}{V(y_{ij})}. \quad (74)$$

(3) Mit Aussagen (2) und (5) von Theorem [Eigenschaften des Modells paralleler Testmessungen](#) gilt

$$R_j = \frac{V(\tau_i)}{V(y_{ij})} = \frac{C(y_{ij}, y_{ik})}{\sqrt{V(y_{ij})}\sqrt{V(y_{ik})}} = \frac{C(y_{ij}, y_{ik})}{\sqrt{V(y_{ij})}\sqrt{V(y_{ik})}} = \rho(y_{ij}, y_{ik}) \quad (75)$$

und dass ebenso

$$R_k = \frac{V(\tau_i)}{V(y_{ik})} = \frac{C(y_{ij}, y_{ik})}{\sqrt{V(y_{ik})}\sqrt{V(y_{ik})}} = \frac{C(y_{ij}, y_{ik})}{\sqrt{V(y_{ij})}\sqrt{V(y_{ik})}} = \rho(y_{ij}, y_{ik}). \quad (76)$$

Normalverteilungsbeispiel

Wir betrachten den Fall zweier Testmessungen $j = 1, 2$ im Modell paralleler Testmessungen. Für alle $i = 1, \dots, n$ seien in Wahrscheinlichkeitsdichteform, wobei wir der notationellen Einfachheit halber auf das i Subskript verzichten wollen,

$$p(t) = N(t; 0, \sigma_\tau^2) \text{ und } p(y_1|t) := N(y_1; t, \sigma_\varepsilon^2) \text{ und } p(y_2|t) := N(y_2; t, \sigma_\varepsilon^2) \quad (77)$$

Für Person i gibt es also nur eine Zufallsvariable des wahren Werts für alle Testmessungen und die Propensitätsverteilungen unterscheiden sich zwischen Testmessungen nicht.

Dann gilt zunächst mit dem [Theorem zu gemeinsamen Normalverteilungen](#) mit $A := 1$ und $b := 0$

$$p(t)p(y_1|t) = p(t, y_1) = N\left(\begin{pmatrix} t \\ y_1 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \quad (78)$$

und weiterhin mit $A := \begin{pmatrix} 1 & 0 \end{pmatrix}$ und $b := 0$

$$p(t, y_1)p(y_2|t) = p(t, y_1, y_2) = N\left(\begin{pmatrix} t \\ y_1 \\ y_2 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \quad (79)$$

Normalverteilungsbeispiel (fortgeführt)

Damit gilt dann durch Ablesen am Kovarianzmatrixparameter von $p(t, y_1, y_2)$, dass

$$\rho(y_1, y_2) = \frac{\mathbb{C}(y_1, y_2)}{\sqrt{\mathbb{V}(y_1)}\sqrt{\mathbb{V}(y_2)}} = \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2}\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2}} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2} \quad (80)$$

Insbesondere gilt also auch für $j = 1, 2$

$$R_j = \rho(y_j, \tau)^2 = \left(\frac{\mathbb{C}(y_j, \tau)}{\mathbb{S}(y_j)\mathbb{S}(\tau)} \right)^2 = \frac{\mathbb{C}(y_j, \tau)^2}{\mathbb{V}(y_j)\mathbb{V}(\tau)} = \frac{(\sigma_\tau^2)^2}{(\sigma_\tau^2 + \sigma_\varepsilon^2)\sigma_\tau^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2} = \rho(y_1, y_2). \quad (81)$$

Gilt also beispielsweise $\sigma_\tau^2 := 1.0$ und $\sigma_\varepsilon^2 := 0.2$, so ergibt sich für die Paralleltestreliabilität

$$R_j = \rho(y_j, \tau)^2 = \rho(y_1, y_2) = \frac{1.0}{1.0 + 0.2} \approx 0.83. \quad (82)$$

Man beachte, dass es sich hierbei um eine auf Grundlage der Modellparameter bestimmte Größe, nicht um eine basierend auf beobachteten Daten geschätzte Größe handelt.

Definition (Paralleltestreliabilitätsschätzer)

Gegeben sei das Modell paralleler Testmessungen für n Testpersonen und zwei Testmessungen $j = 1, 2$,

$$\mathbb{P}(\tau_1, y_{11}, y_{12}, \dots, \tau_n, y_{n1}, y_{n2}) = \prod_{i=1}^n \mathbb{P}(\tau_i) \mathbb{P}(y_{i1} | \tau_i) \mathbb{P}(y_{i2} | \tau_i) \quad (83)$$

Dann wird der mit den Stichprobenmitteln

$$\bar{y}_1 := \frac{1}{n} \sum_{i=1}^n y_{i1} \text{ und } \bar{y}_2 := \frac{1}{n} \sum_{i=1}^n y_{i2} \quad (84)$$

definierte Stichprobenkorrelationskoeffizient

$$r_{12} := \frac{\frac{1}{n-1} \sum_{i=1}^n (y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{i1} - \bar{y}_1)^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{i2} - \bar{y}_2)^2}} \quad (85)$$

Paralleltestreliabilitätsschätzer genannt und häufig mit r_{tt} bezeichnet.

Bemerkungen

- **R** bietet mit der `cor()` Funktion eine Möglichkeit, Stichprobenkorrelationskoeffizienten zu berechnen.

Normalverteilungsbeispiel (fortgeführt)

Simulation des Normalverteilungsbeispiels für $n := 30$

$p(t_i) = N(t_i; 0, \sigma_\tau^2)$ und $p(y_{ij}|t_i) := N(y_{ij}; t_i, \sigma_\varepsilon^2)$ für $j = 1, 2$ mit $\sigma_\tau^2 := 1.0$ und $\sigma_\varepsilon^2 := 0.2$ für $i = 1, \dots, n$.

```
n          = 30                # Personenanzahl
m          = 2                 # Testmessungsanzahl
sigsqr_tau = 1                 # wahrer Wert Varianz
sigsqr_eps = .2                # beobachteter Wert-Varianz
R_12       = sigsqr_tau/(sigsqr_tau+sigsqr_eps) # Paralleltestreliabilität
T          = matrix(rep(NaN, n), nrow = n)      # wahrer Wert Array
Y          = matrix(rep(NaN, n*m), nrow = n)    # beobachteter Wert Array
E          = matrix(rep(NaN, n*m), nrow = n)    # Messfehler Array
for(i in 1:n){
  T[i] = rnorm(1,1,sqrt(sigsqr_tau))             # Personeniterationen
  for(j in 1:m){
    Y[i,j] = rnorm(1,T[i],sqrt(sigsqr_eps))      # wahrer Wert Realisierung für j = 1,2
    E[i,j] = Y[i,j] - T[i]}                     # Testmessungsiterationen
  }                                              # beobachteter Wert Realisierung f
  E[i,j] = Y[i,j] - T[i]}                     # Messfehler Realisierung
r_12     = cor(Y[,1],Y[,2])                    # Reliabilitätsschätzer
```

Paralleltestreliabilität R_12 : 0.8333

Paralleltestreliabilitätsschätzer r_12 : 0.8153

Praxis der Paralleltestreliabilitätsschätzung

Die Items einer Paralleltestform müssen gleichzeitig anders und ähnlich zum Originaltest sein.

Je ähnlicher die Items, desto größer sollte der zeitliche Testabstand sein, um Lerneffekte zu vermindern.

Je unähnlicher die Items, desto kürzer sollte der zeitliche Testabstand sein, um systematische Einflüsse zu verringern.

Bei Leistungstests ist dies zum Teil möglich

- Für manche Aufmerksamkeits- und Schuleignungstests existieren Parallelförmn
- Für manche Subtests von Intelligenztests existieren Parallelförmn

Praktisch werden Teststichprobenmittel, -standardabweichungen, und -korrelationen bestimmt

Ähneln sich Stichprobenmittel und -varianzen bei hoher Korrelation, so schließt man auf Parallelität

Bei Persönlichkeitsfragebögen oder klinischen Tests gestaltet sich dies eher schwierig

- Meist sind diese Tests gerade durch ihre speziellen Items charakterisiert
- Es müssten große Pools ähnlicher Items existieren

Moosbrugger & Kelava (2012), Kapitel 6, Pospeschill (2022), Kapitel 6

Praxis der Retestreliabilitätsschätzung

Derselbe Test wird an derselben Stichprobe zweimal durchgeführt.

Die Korrelation der Messwerte über beide Darbietungen wird berechnet und als r_{tt} bezeichnet.

Es wird vorausgesetzt, dass die Annahmen des Modells paralleler Testmessungen gelten.

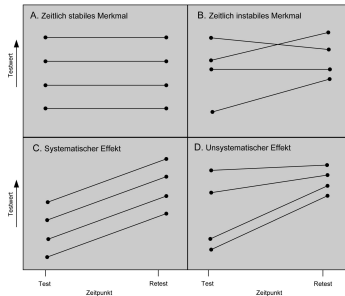
- Der wahre Wert einer Testperson bleibt über die Testmessungen hinweg konstant
- Die bedingten Varianzen der beobachteten Werte bleiben über Testwiederholungen konstant

Mögliche Änderungen der wahren Werte im Zeitverlauf

- Unsystematische Veränderungen oder situative Faktoren, die über Testpersonen unterschiedlich ausgeprägt sind, wie z.B. interindividuell unterschiedliche Übungseffekte, Tagesformen, instabile Merkmale oder Erinnerungseffekte verringern die Retestreliabilität. Kurze Test-Retest-Intervalle minimieren dabei unsystematische Merkmalsveränderungen, aber maximieren Erinnerungseffekte und vice versa.
- Systematische Veränderungen oder situative Faktoren, die alle Testpersonen gleich betreffen, ändern die Retestreliabilität nicht. Im Modell der multiplen Testmessungen entspricht dies den hier nicht weiter vertieften wesentlich τ -äquivalenten Testmessungen.

Moosbrugger & Kelava (2012), Kapitel 6, Pospeschill (2022), Kapitel 6

Potentielle Dynamiken wahrer Werte von $n = 4$ Testpersonen über Testmessungen



aus Pospeschill (2022), Kapitel 6

- | | |
|--|---------------------------------|
| A Identische wahre Werte bei einem stabilen Merkmal | ⇒ Hohe Reliabilität |
| B Variierende wahre Werte bei einem instabilen Merkmal | ⇒ Verringerung der Reliabilität |
| C Identische linear-affine Transformation des wahren Werts | ⇒ Hohe Reliabilität |
| D Variable positiv linear-affine Transformation des wahren Werts | ⇒ Verringerung der Reliabilität |

Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

Motivation

BDI-II Interne Konsistenz (Hautzinger et al. (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

Vorbemerkungen

y_{ij} bezeichnete bisher die Zufallsvariable zur Modellierung des beobachteten Werts der j ten Testmessung der i ten Testperson, wobei wir m Testmessungen $j = 1, \dots, m$ und n Testpersonen $i = 1, \dots, n$ angenommen haben.

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

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

y_{ij} bezeichnet also nun die Zufallsvariable zur Modellierung des beobachteten Werts des j ten Items der i ten Testperson in einem Test, wobei wir weiterhin m Items $j = 1, \dots, m$ und n Testpersonen $i = 1, \dots, n$ annehmen.

Insbesondere gehen wir davon aus, dass für jede Testperson eine Summe von beobachteten Werten durch Summation über die Items eines Test gebildet wird. Die Zufallsvariable zur Modellierung dieser Summe von beobachteten Werten bezeichnen wir mit

$$y_i := \sum_{j=1}^m y_{ij} \quad (86)$$

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

Definition (m -Komponententestmodell)

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

- $y_i := \sum_{j=1}^m y_{ij}$ die *Summe der beobachteten Werte* und
- $\tau_i := \sum_{j=1}^m \tau_{ij}$ die *Summe der wahren Werte*.

Dann heißt die gemeinsame Verteilung der y_i und τ_i für $i = 1, \dots, n$ mit der Faktorisierungseigenschaft

$$\mathbb{P}(\tau_1, \dots, \tau_n, y_1, \dots, y_n) := \prod_{i=1}^n \mathbb{P}(\tau_i, y_i) = \prod_{i=1}^n \mathbb{P}(y_i | \tau_i) \mathbb{P}(\tau_i) \quad (87)$$

das m -Komponententestmodell, wenn gilt dass

$$\mathbb{P}(\tau_1, y_1) = \dots = \mathbb{P}(\tau_n, y_n). \quad (88)$$

Bemerkung

- Die m Komponenten eines Tests entsprechen seinen m Items.

Definition (Reliabilitäten von m -Komponententestmodellen)

Gegeben sei ein m -Komponententestmodell.

(1) Für $j = 1, \dots, m$ ist die Reliabilität der j ten Testmessung definiert als

$$R_j := \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} \text{ für ein beliebiges } 1 \leq i \leq n. \quad (89)$$

(2) Die Gesamtreliabilität des m -Komponententestmodells ist definiert

$$R := \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_i)} \text{ für ein beliebiges } 1 \leq i \leq n. \quad (90)$$

Bemerkung

- Die Definitionen nutzen die erste Umformung der im vorherigen Abschnitt genutzten Definition.
- In R_j gehen die Varianzen des j ten wahren Werts τ_{ij} und j ten beobachteten Werts y_{ij} ein.
- In R gehen die Varianzen der Summe τ_i der wahren Werte und der Summe y_i der beobachteten Werte ein.
- Wenn das m -Komponententestmodell einen psychologischen Test bestehend aus m Items modelliert, dann wird R_j auch *Itemreliabilität* des j ten Items und R *Reliabilität des Tests* genannt.

Cronbach's α

Grundlage der Bestimmung von Cronbach's α ist die Idee der *Testhalbierungs-Reliabilität*

- Besteht ein Test aus vielen Items, so kann eine Testhälfte als Paralleltest der anderen angesehen werden.
- Die Korrelation der Testhälften ist dann ein Maß für die Paralleltestreliabilität.

Die resultierende Testhälften-Korrelation ist aber nur die Reliabilität eines Tests halber Länge

Grundsätzlich erhöht sich die Reliabilität eines Test paralleler Items mit zunehmender Testlänge

- Die Varianz der Messfehler erhöht sich linear, da Messfehler nach Annahme unabhängig sind.
- Die Varianz der wahren Werte erhöht sich quadratisch, da wahre Werte nach Annahme kovariieren.

Die Splithalf-Korrelation muss also auf die volle Testlänge aufgewertet werden.

- Dies ist mit der von Spearman (1910) und Brown (1910) entwickelten Formel möglich

Man kann einen Test natürlich auch in mehr als zwei Teile teilen und alle Paralleltestkorrelationen bestimmen.

- Cronbach's α bildet die Situation ab, in der alle Items als Paralleltests betrachtet werden.

Wir betrachten in der Folge die Spearman-Brown Formel und ihre Eigenschaften als Spezialfall von Cronbach's α .

Definition (Cronbach's α)

Gegeben sei ein m -Komponententestmodell. Dann heißt

$$\alpha := \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_{ij})}{\mathbb{V}(y_i)} \right) \text{ für beliebiges } 1 \leq i \leq n \quad (91)$$

Cronbach's α oder Koeffizient α .

Bemerkungen

- $\mathbb{V}(y_{ij})$ ist die Varianz der beobachteten Werte für Testpersonen $i = 1, \dots, n$ und Items $j = 1, \dots, m$.
- $\mathbb{V}(y_i)$ ist die Varianz der Summe der beobachteten Werte für Testpersonen $i = 1, \dots, n$.
- Die Formel wurde durch Cronbach (1951) als *Coefficient α* popularisiert
- Frühere Anwendungen finden sich zum Beispiel in Hoyt (1941) und Guttman (1945)

Bemerkungen (fortgesetzt)

- Da nach Annahme des Modells multipler Testmessungen die Verteilungen der beobachteten Werte über Personen identisch sind, wird Cronbach's α häufig auch geschrieben als

$$\alpha := \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)} \right) \quad (92)$$

- Zum Verständnis von α unterstellen wir im Folgenden, dass die Anzahl der Items hoch, also $\frac{m}{m-1} \approx 1$ ist.
- Nach Theorem [Varianz einer Linearkombination](#) gilt dann allgemein

$$\mathbb{V}(y) = \sum_{j=1}^m \mathbb{V}(y_j) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \mathbb{C}(y_j, y_k) \quad (93)$$

- Sind die beobachteten Werte in diesem Fall (marginal) unabhängig, so gilt

$$\mathbb{C}(y_j, y_k) = 0 \text{ für alle } 1 \leq j, k \leq m \quad (94)$$

- Bei großem m und Unabhängigkeit der beobachteten Werte gilt also

$$\alpha \approx 1 \cdot \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)} \right) = \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\sum_{j=1}^m \mathbb{V}(y_j)} \right) = 0 \quad (95)$$

- Induziert ein konstanter wahrer Wert andererseits marginale Kovarianzen $\mathbb{C}(y_j, y_k) > 0$, so folgt

$$\mathbb{V}(y) = \sum_{j=1}^m \mathbb{V}(y_j) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \mathbb{C}(y_j, y_k) \gg \sum_{j=1}^m \mathbb{V}(y_j) \Rightarrow \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)} \approx 0 \Rightarrow \alpha \approx 1 \quad (96)$$

Theorem (Cronbach's α und Gesamtreliabilität)

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

$$\alpha \leq R \quad (97)$$

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

Bemerkungen

- Cronbach's α ist eine untere Grenze für die Gesamtreliabilität eines m -Komponententestmodells.
- Die Gesamtreliabilität eines m -Komponententestmodells ist mindestens so groß wie Cronbach's α .
- Die Gesamtreliabilität eines m -Komponententestmodells kann größer als Cronbach's α sein.
- Für parallele Testmessungen (Items) ist die Gesamtreliabilität eines m -Komponententestmodells gleich α .
- Der Beweis des Theorems geht unter anderem auf Novick & Lewis (1967) zurück.

Beweis

Zur Vereinfachung der Notation verzichten wir auf die explizite Auszeichnung der Testpersonen $i = 1, \dots, n$ und setzen

$$y_j := y_{ij}, \tau_j := \tau_{ij}, y := \sum_{j=1}^m y_j \text{ und } \tau := \sum_{j=1}^m \tau_j, \quad (98)$$

sowie

$$R := \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} \text{ und } \alpha := \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)} \right). \quad (99)$$

Wir gehen in vier Schritten vor.

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

$$\sum_{j=1}^m x_j = \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m (x_j + x_k). \quad (100)$$

Beweis

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

$$\begin{aligned}\frac{1}{3} \sum_{j=1}^3 \sum_{k=j+1}^4 (x_j + x_k) &= \frac{1}{3} \left(\sum_{k=1+1}^4 (x_1 + x_k) + \sum_{k=2+1}^4 (x_2 + x_k) + \sum_{k=3+1}^4 (x_3 + x_k) \right) \\&= \frac{1}{3} \left(\sum_{k=2}^4 (x_1 + x_k) + \sum_{k=3}^4 (x_2 + x_k) + \sum_{k=4}^4 (x_3 + x_k) \right) \\&= \frac{1}{3} ((x_1 + x_2) + (x_1 + x_3) + (x_1 + x_4) + (x_2 + x_3) + (x_2 + x_4) + (x_3 + x_4)) \\&= \frac{1}{3} ((x_1 + x_1 + x_1) + (x_2 + x_2 + x_2) + (x_3 + x_3 + x_3) + (x_4 + x_4 + x_4)) \\&= \frac{1}{3} (3x_1 + 3x_2 + 3x_3 + 3x_4) \\&= x_1 + x_2 + x_3 + x_4 \\&= \sum_{j=1}^4 x_j\end{aligned} \tag{101}$$

Beweis (fortgeführt)

(2) (*Kovarianzungleichung für den wahren Wert*) Wir leiten nun im Modell multipler Testmessungen eine Ungleichung her. Dazu betrachten wir im Modell multipler Testmessungen die Varianz der Differenz zweier wahrer Werte τ_j und τ_k . Mit Theorem [Nichtnegativität der Varianz](#) und Theorem [Varianzen spezieller Linearkombinationen](#) ergibt sich

$$\mathbb{V}(\tau_j - \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) - 2\mathbb{C}(\tau_j, \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) \geq 2\mathbb{C}(\tau_j, \tau_k) \quad (102)$$

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

$$\mathbb{V}(\tau_j - \tau_k) = \mathbb{V}(f_j(\tau_1) - f_k(\tau_1)) = \mathbb{V}(\tau_1 - \tau_1) = \mathbb{V}(0) = 0. \quad (103)$$

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

(3) (*Varianzungleichung für die Summe wahrer Werte*) Wir betrachten nun die Varianz der Summe der wahren Werte im m -Komponententestmodell. Mit Theorem [Varianz einer Linearkombination](#) der Summendarstellung, aus (1) und der Kovarianzungleichung für den wahren Wert aus (2) ergibt sich zunächst

Beweis (fortgeführt)

$$\begin{aligned}\mathbb{V}(\tau) &= \mathbb{V}\left(\sum_{j=1}^m \tau_j\right) \\&= \sum_{j=1}^m \mathbb{V}(\tau_j) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \mathbb{C}(\tau_j, \tau_k) \\&= \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m (\mathbb{V}(\tau_j) + \mathbb{V}(\tau_k)) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \mathbb{C}(\tau_j, \tau_k) \\&\geq \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k) + \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k) \\&= \left(\frac{1}{m-1} + 1\right) \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k) \\&= \left(\frac{1}{m-1} + \frac{m-1}{m-1}\right) \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k) \\&= \frac{1+m-1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k) \\&= \frac{m}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k)\end{aligned}\tag{104}$$

Beweis (fortgeführt)

Mit Aussage (5) von Theorem [Eigenschaften bezüglich zweier Testmessungen](#) und wiederum mit Theorem [Varianz einer Linearkombination](#) gilt dann

$$\begin{aligned}\mathbb{V}(\tau) &\geq \frac{m}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(\tau_j, \tau_k) \\ &= \frac{m}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^m 2\mathbb{C}(y_j, y_k) \\ &= \frac{m}{m-1} \left(\mathbb{V} \left(\sum_{j=1}^m y_j \right) - \sum_{j=1}^m \mathbb{V}(y_j) \right) \\ &= \frac{m}{m-1} \left(\mathbb{V}(y) - \sum_{j=1}^m \mathbb{V}(y_j) \right)\end{aligned}\tag{105}$$

Beweis (fortgeführt)

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

$$\begin{aligned} R &= \frac{V(\tau)}{V(y)} \\ &\geq \frac{\frac{m}{m-1} \left(V(y) - \sum_{j=1}^m V(y_j) \right)}{V(y)} \\ &= \frac{m}{m-1} \left(\frac{V(y)}{V(y)} - \frac{\sum_{j=1}^m V(y_j)}{V(y)} \right) \\ &= \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m V(y_j)}{V(y)} \right) \\ &=: \alpha. \end{aligned} \tag{106}$$

□

Definition (Cronbach's α Schätzer)

Gegeben sei das Modell m -Komponententestmodell mit beobachteten Werten y_{ij} für $i = 1, \dots, n$ Testpersonen und $j = 1, \dots, m$ Testkomponenten (Items). Dann ist ein Schätzer für die Varianz des j ten Items gegeben durch

$$S_j^2 := \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 \text{ mit } \bar{y}_j := \frac{1}{n} \sum_{i=1}^n y_{ij} \quad (107)$$

und ein Schätzer für die Varianz der Testsummenscores ist gegeben durch

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \text{ mit } y_i := \sum_{j=1}^m y_{ij} \text{ und } \bar{y} := \frac{1}{n} \sum_{i=1}^n y_i \quad (108)$$

Ein Schätzer für α ergibt sich entsprechend zu

$$\hat{\alpha} = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m S_j^2}{S^2} \right). \quad (109)$$

Bemerkung

- Zur frequentistischen Verteilungstheorie dieses Schätzers, siehe z.B. Kristof (1963), Feldt (1965), Feldt et al. (1987), Van Zyl et al. (2000), Yuan & Bentler (2002).

Simulation im Modell paralleler Testmessungen

$$n := 30, m := 21 \text{ mit } \mathbb{P}(\tau_i) := N(1, 1), \mathbb{P}(y_{ij}|\tau_i) := N(\tau_i, 4) \text{ für } i = 1, \dots, n, j = 1, \dots, m. \quad (110)$$

```
library(psych)                                # R Paket zur Testanalyse
set.seed(0)                                  # Reproduzierbarkeit
n  = 30                                       # Personenanzahl
m  = 21                                       # Itemanzahl
mu  = 1                                       # Wahrer Wert Erwartungswertparameter
T  = matrix(rep(NaN, n), nrow = n)          # Wahrer Wert Array
Y  = matrix(rep(NaN, n*m), nrow = n)        # Beobachteter Wert Array
for(i in 1:n){                               # Iteration über Testpersonen
  T[i] = rnorm(2,mu,sqrt(1))                 # Wahrer Wert Realisierung
  for(j in 1:m){                             # Iteration über Items
    Y[i,j] = rnorm(1,T[i],sqrt(4))}}         # Beobachteter Wert Realisierung
vsi  = var(apply(Y,1,sum))                   # Stichprobenvarianz der beobachteter Wert-Summern
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

Theorem (Spearman-Brown-Formel)

Gegeben sei ein m -Komponententestmodell mit parallelen Testmessungen, Reliabilität R und Komponentenreliabilitäten $R_1, \dots, R_m$. Dann gilt

$$R = \frac{mR_1}{1 + (m-1)R_1} \quad (111)$$

Bemerkungen

- Die Testhalbierungsreliabilität ergibt sich als der Spezialfall $m = 2$.
- Sei dazu R_1 die durch $\rho(y_{i1}, y_{i2})$ gegebene Paralleltestreliabilität. Für die Gesamttestreliabilität ergibt sich dann

$$R = \frac{2 \cdot R_1}{1 + (2-1)R_1} = \frac{2\rho(y_{i1}, y_{i2})}{1 + \rho(y_{i1}, y_{i2})} \quad (112)$$

- Für $R_1 = 0.7$ ergibt sich beispielsweise

$$R = \frac{2 \cdot 0.7}{1 + 0.7} \approx 0.82 \quad (113)$$

- In der Praxis unterscheidet man folgende Methoden zur Testhalbierung:
- Odd-Even: Eine Testhälfte besteht aus den ungerade, die andere aus den gerade nummerierten Items
- Zeitpartitionierung: Der Test wird in zwei Abschnitte geteilt, deren Bearbeitung die gleiche Zeit beansprucht.
- Itemzwillinge: Es werden Itemzwillinge konstruiert und zufällig einer der Testhälften zugeordnet.

Beweis

Wir halten zunächst fest, dass nach Theorem [Cronbach's \$\alpha\$ und Gesamtreliabilität](#) und durch Annahme paralleler Testmessungen gilt, dass

$$R = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)} \right) \quad (114)$$

(1) (*Summe der Varianzen der beobachteten Werte bei Parallelität*) Nach Theorem [Eigenschaften des Modells paralleler Testmessungen](#) sind die Varianzen aller beobachteten Werte über Testmessungen bei parallelen Testmessungen identisch. Es gilt also

$$\sum_{j=1}^m \mathbb{V}(y_j) = \sum_{j=1}^m \mathbb{V}(y_1) = m\mathbb{V}(y_1) \quad (115)$$

Beweis (fortgeführt)

(2) (Varianz der Summe der beobachteten Werte bei Parallelität)

Mit Theorem [Varianz einer Linearkombination](#) gilt

$$\begin{aligned}\mathbb{V}(y) &= \mathbb{V}\left(\sum_{j=1}^m y_j\right) \\&= \sum_{j=1}^m \mathbb{V}(y_j) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \mathbb{C}(y_j, y_k) \\&\stackrel{(a)}{=} \sum_{j=1}^m \mathbb{V}(y_1) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \mathbb{C}(y_1, y_2) \\&= m\mathbb{V}(y_1) + 2\mathbb{C}(y_1, y_2) \sum_{j=1}^{m-1} \sum_{k=j+1}^m 1 \\&\stackrel{(b)}{=} m\mathbb{V}(y_1) + 2\mathbb{C}(y_1, y_2) \frac{m(m-1)}{2} \\&= m\mathbb{V}(y_1) + m(m-1)\rho(y_1, y_2)\sqrt{\mathbb{V}(y_1)}\sqrt{\mathbb{V}(y_2)} \\&\stackrel{(c)}{=} m\mathbb{V}(y_1) + m(m-1)\rho(y_1, y_2)\sqrt{\mathbb{V}(y_1)}\sqrt{\mathbb{V}(y_1)} \\&\stackrel{(d)}{=} m\mathbb{V}(y_1) + m(m-1)R_1\mathbb{V}(y_1) \\&= m\mathbb{V}(y_1)(1 + (m-1)R_1)\end{aligned}\tag{116}$$

Beweis (fortgeführt)

Dabei gilt

- $\stackrel{(a)}{=}$ mit Aussage (5) von Theorem [Eigenschaften des Modells paralleler Testmessungen](#), dass

$$\mathbb{C}(y_j, y_k) = \mathbb{V}(\tau) \text{ für alle } j, k \text{ mit } 1 \leq j, k \leq m \quad (117)$$

also, dass die Kovarianzen der beobachteten Werte beliebiger Paare paralleler Testmessungen identisch und damit insbesondere identisch zu $\mathbb{C}(y_1, y_2)$ sind,

- $\stackrel{(b)}{=}$ mit der unter (4) per Induktionsbeweis gezeigten Identität

$$\sum_{j=1}^{m-1} \sum_{k=j+1}^m 1 = \frac{m(m-1)}{2} \quad (118)$$

- $\stackrel{(c)}{=}$ mit Aussage (2) von Theorem [Eigenschaften des Modells paralleler Testmessungen](#),
- $\stackrel{(d)}{=}$ mit Aussage (3) von Theorem [Parallelttestreliabilität](#)

Beweis (fortgeführt)

(3) (Einsetzen in Cronbach's α)

Einsetzen ergibt dann

$$\begin{aligned} R &= \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)} \right) \\ &= \frac{m}{m-1} \left(1 - \frac{m\mathbb{V}(y_1)}{m\mathbb{V}(y_1)(1+(m-1)R_1)} \right) \\ &= \frac{m}{m-1} \left(1 - \frac{1}{1+(m-1)R_1} \right) \\ &= \frac{m}{m-1} \left(\frac{1+(m-1)R_1}{1+(m-1)R_1} - \frac{1}{1+(m-1)R_1} \right) \\ &= \frac{m}{m-1} \left(\frac{1+(m-1)R_1-1}{1+(m-1)R_1} \right) \\ &= \frac{m}{m-1} \cdot \frac{(m-1)R_1}{1+(m-1)R_1} \\ &= \frac{mR_1}{1+(m-1)R_1}. \end{aligned} \tag{119}$$

Beweis (fortgeführt)

(4) *(Induktionsbeweis der Doppelsummenform)*

Es bleibt zu zeigen, dass

$$S_m := \sum_{j=1}^{m-1} \sum_{k=j+1}^m 1 = \frac{m(m-1)}{2} =: P_m \quad (120)$$

(1) Induktionsanfang: Für $m := 2$ gilt

$$S_2 = \sum_{j=1}^{2-1} \sum_{k=j+1}^2 1 = \sum_{j=1}^1 \sum_{k=j+1}^2 1 = \sum_{k=1+1}^2 1 = \sum_{k=2}^2 1 = 1 = \frac{2(2-1)}{2} = P_2 \quad (121)$$

Zu Demonstrationszwecken betrachten wir auch $m := 3$. Hier gilt

$$S_3 = \sum_{j=1}^{3-1} \sum_{k=j+1}^3 1 = \sum_{j=1}^2 \sum_{k=j+1}^3 1 = \sum_{k=1+1}^3 1 + \sum_{k=2+1}^3 1 = \sum_{k=2}^3 1 + \sum_{k=3}^3 1 = 3 = \frac{3(3-1)}{2} = P_3 \quad (122)$$

(2) Induktionsschritt: Es gelte nun für ein $m \geq 2$

$$S_m = \sum_{j=1}^{m-1} \sum_{k=j+1}^m 1 = \frac{m(m-1)}{2} = P_m \quad (123)$$

Beweis (fortgeführt)

Dann gilt nach Definition für $m + 1$

$$\begin{aligned} S_{m+1} &= \sum_{j=1}^{m+1-1} \sum_{k=j+1}^{m+1} 1 \\ &= \sum_{j=1}^m \sum_{k=j+1}^{m+1} 1 \\ &= \sum_{j=1}^{m-1} \sum_{k=j+1}^{m+1} 1 + \sum_{k=m+1}^{m+1} 1 \\ &= \left(\sum_{j=1}^{m-1} \sum_{k=j+1}^{m+1} 1 \right) + 1 \\ &= \left(\sum_{j=1}^{m-1} \left(\sum_{k=j+1}^m 1 + 1 \right) \right) + 1 \\ &= \left(\sum_{j=1}^{m-1} \sum_{k=j+1}^m 1 \right) + \left(\sum_{j=1}^{m-1} 1 \right) + 1 \\ &= \left(\sum_{j=1}^{m-1} \sum_{k=j+1}^m 1 \right) + (m-1) + 1 \\ &= S_m + m \end{aligned} \tag{124}$$

Beweis (fortgeführt)

Dann aber gilt nach der Induktionsschrittannahme

$$\begin{aligned} S_{m+1} &= S_m + m \\ &= P_m + m \\ &= \frac{m(m-1)}{2} + m \\ &= \frac{m(m-1)}{2} + \frac{2m}{2} \\ &= \frac{m(m-1) + 2m}{2} \\ &= \frac{m^2 - m + 2m}{2} \\ &= \frac{m^2 + m}{2} \\ &= \frac{m(m+1)}{2} \\ &= \frac{(m+1)((m+1)-1)}{2} \\ &= P_{m+1} \end{aligned} \tag{125}$$

Normalverteilungsbeispiel

Wir betrachten den Fall eines m -Komponententestmodells mit m parallelen Testmessungen. Für alle $i = 1, \dots, n$ seien in Wahrscheinlichkeitsdichteform, wobei wir wiederum auf das Subskript i verzichten wollen

$$p(t_1) := N(t_1; 0, \sigma_\tau^2), t_2 := t_1, \dots, t_m := t_1 \quad (126)$$

sowie

$$p(y_1|t_1) := N(y_1; t_1, \sigma_\varepsilon^2), \dots, p(y_m|t_1) := N(y_m; t_1, \sigma_\varepsilon^2) \quad (127)$$

Per definition gilt im m -Komponententest mit parallelen Testmessungen für die Gesamtreliabilität

$$R := \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} = \frac{\mathbb{V}\left(\sum_{j=1}^m \tau_i\right)}{\mathbb{V}\left(\sum_{j=1}^m y_i\right)} \quad (128)$$

Für den Zähler halten wir fest, dass mit der Identität der wahren Werte im Modell paralleler Testmessungen gilt

$$\mathbb{V}\left(\sum_{j=1}^m \tau_i\right) = \mathbb{V}\left(\sum_{j=1}^m \tau_1\right) = \mathbb{V}(m\tau_1) = m^2 \mathbb{V}(\tau_1) = m^2 \sigma_\tau^2 \quad (129)$$

Normalverteilungsbeispiel (fortgeführt)

Für den Nenner betrachten wir die Wahrscheinlichkeitsdichtefunktion der gemeinsamen Verteilung von $\tau_1, y_1, \dots, y_m$, die in Generalisierung des im Abschnitt Reliabilität betrachteten Normalverteilungsbeispiels die folgende Form hat:

$$p(t_1, y_1, \dots, y_m) = N \left(\begin{pmatrix} t_1 \\ y_1 \\ \vdots \\ y_m \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 & \dots & \sigma_\tau^2 \\ \sigma_\tau^2 + \sigma_\varepsilon^2 & \sigma_\tau^2 & \dots & \sigma_\tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\tau^2 & \sigma_\tau^2 & \dots & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix} \right) \quad (130)$$

Um nun die Varianz der Summe der beobachteten Werte zu bestimmen, bestimmen wir mithilfe des [Theorems zur linear-affinen Transformation eines normalverteilten Zufallsvektors](#) die Kovarianzmatrix des durch die Matrix $A := \begin{pmatrix} 0 & 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{1 \times (m+1)}$ transformierten Zufallsvektors $(t, y_1, \dots, y_m)^T$, also die Kovarianzmatrix von

$$\sum_{j=1}^m y_j = \begin{pmatrix} 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ y_1 \\ \vdots \\ y_m \end{pmatrix} \quad (131)$$

Interne Konsistenz

Normalverteilungsbeispiel (fortgeführt)

Es ergibt sich

$$\begin{pmatrix} 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \sigma_{\tau}^2 & \sigma_{\tau}^2 & \dots & \sigma_{\tau}^2 \\ \sigma_{\tau}^2 & \sigma_{\tau}^2 + \sigma_{\varepsilon}^2 & \dots & \sigma_{\tau}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\tau}^2 & \sigma_{\tau}^2 & \dots & \sigma_{\tau}^2 + \sigma_{\varepsilon}^2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} m\sigma_{\tau}^2 & m\sigma_{\tau}^2 + \sigma_{\varepsilon}^2 & \dots & m\sigma_{\tau}^2 + \sigma_{\varepsilon}^2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ = m(m\sigma_{\tau}^2 + \sigma_{\varepsilon}^2) \\ = m^2\sigma_{\tau}^2 + m\sigma_{\varepsilon}^2$$

Für die Gesamtreliabilität ergibt sich folglich

$$R = \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} = \frac{\mathbb{V}\left(\sum_{j=1}^m \tau_i\right)}{\mathbb{V}\left(\sum_{j=1}^m y_i\right)} = \frac{m^2\sigma_{\tau}^2}{m^2\sigma_{\tau}^2 + m\sigma_{\varepsilon}^2} \quad (132)$$

Gilt also beispielsweise $\sigma_{\tau}^2 := 1.0$ und $\sigma_{\varepsilon}^2 := 0.2$, so ergibt sich für R bei $m := 21$ parallelen Testmessungen

$$R = \frac{(21)^2 \cdot 1.0}{(21)^2 \cdot 1.0 + 21 \cdot 0.2} = \frac{441}{441 + 4.2} = 0.990566 \quad (133)$$

Basierend auf der im Abschnitt Reliabilität bestimmten Itemreliabilität $R_1 = \frac{1.0}{1.2}$ ergibt sich mithilfe der Spearman-Brown-Formel dasselbe Ergebnis

$$R = \frac{mR_1}{1 + (m-1)R_1} = \frac{21 \cdot \frac{1.0}{1.2}}{1 + 20 \cdot \frac{1.0}{1.2}} = 0.990566. \quad (134)$$

Normalverteilungsbeispiel

```
n      = 1e1      # Anzahl Testpersonen
m      = 21      # Anzahl Testmessungen
k      = 2       # Anzahl Testwiederholungen
sigsqr_tau = 1    # Varianz der wahren Werte
sigsqr_eps = 0.2  # Varianz der Messfehler
R_1     = sigsqr_tau/(sigsqr_tau + sigsqr_eps) # Itemreliabilität
R       = ((m^2)*sigsqr_tau)/(((m^2)*sigsqr_tau + m*sigsqr_eps)) # Gesamtreliabilität nach Beispiel
SBF     = (m*R_1)/(1 + (m-1)*R_1)           # Gesamtreliabilität nach SB-Formel
T       = matrix(rep(NaN,n), nrow = n)      # Initialisierung wahre Werte
Y       = array(NaN, dim = c(n, m, k))      # Initialisierung beobachtete Werte
for(i in 1:n){
  T[i] = rnorm(1,1,sqrt(sigsqr_tau))         # Iteration über Testpersonen
  for(t in 1:k){
    for(j in 1:m){
      Y[i,j,t] = rnorm(1,T[i], sqrt(sigsqr_eps))}}} # Realisierung wahrer Werte pro Testperson
R_tt    = cor(apply(Y, c(1, 3), sum))[1,2]   # Iteration über parallele Testwiederholungen
                                              # Iteration über parallele Testmessungen
                                              # Realisierung beobachteter Werte
                                              # Korrelation der Summenscores
```

```
Itemereliabilität      : 0.8333333
Gesamtreliabilität nach Normalverteilungsbeispiel : 0.990566
Gesamtreliabilität nach Spearman-Brown-Formel    : 0.990566
Korrelation der Summenscores über Paralleltests   : 0.9974069
```

Grundlagen

Das Modell multipler Testmessungen

Das Modell paralleler Testmessungen

Reliabilität

Interne Konsistenz

Selbstkontrollfragen

1. Skizzieren Sie die Klassische Testtheorie und Ihre Bedeutung für die Fragebogendatenanalyse.
2. Geben Sie die Definition eines bedingten Erwartungswerts wieder.
3. Geben Sie die Definitionen von wahrem Wert, beobachteten Wert und Messfehler wieder.
4. Erläutern Sie den Begriff der Propensitätsverteilung im Kontext der Klassischen Testtheorie.
5. Geben Sie die Definition des Modells multipler Testmessungen wieder.
6. Erläutern Sie das Modell multipler Testmessungen.
7. Geben Sie die Definition des Vereinfachten Modells der Klassischen Testtheorie wieder.
8. Erläutern Sie das Vereinfachte Modell der Klassischen Testtheorie.
9. Geben Sie das Theorem zu den Eigenschaften bezüglich einer Testmessung wieder.
10. Erläutern Sie das Theorem zu den Eigenschaften bezüglich einer Testmessung.
11. Geben Sie das Theorem zu den Eigenschaften bezüglich zweier Testmessungen wieder.
12. Erläutern Sie das Theorem zu den Eigenschaften bezüglich zweier Testmessungen.
13. Erläutern Sie den Begriff der parallelen Testmessungen.
14. Erläutern Sie die Bedeutungen von

$$\tau_{ij} = \text{id}_{\mathbb{R}}(\tau_{i1}) \text{ und } t_{ij} = \text{id}_{\mathbb{R}}(t_{i1}) \text{ für } i = 1, \dots, n \text{ und } j = 2, \dots, m \quad (135)$$

im Kontext paralleler Testmessungen.

15. Geben Sie die Definition des Modells paralleler Testmessungen wieder.
16. Erläutern Sie das Modell paralleler Testmessungen.
17. Geben Sie das Theorem zu den Eigenschaften des Modells paralleler Testmessungen wieder.
18. Erläutern Sie das Theorem zu den Eigenschaften des Modells paralleler Testmessungen.
19. Geben Sie die Definition der Reliabilität einer Testmessung wieder.
20. Geben Sie das Theorem zu den Eigenschaften der Reliabilität einer Testmessung wieder.

21. Erläutern Sie die drei äquivalenten Formen der Reliabilität einer Testmessung im Modell multipler Testmessungen.
22. Geben Sie die Definition der Reliabilität einer Paralleltestmessung wieder.
23. Geben Sie das Theorem zur Paralleltestreliabilität wieder.
24. Warum ist die Annahme paralleler Testmessungen für die empirische Messung der Reliabilität zentral?
25. Geben Sie die Definition des Paralleltestreliabilitätsschätzers wieder.
26. Erläutern Sie das praktische Vorgehen zur Durchführung einer Paralleltestreliabilitätsschätzung.
27. Erläutern Sie das praktische Vorgehen zur Durchführung einer Retestreliabilitätsschätzung.
28. Diskutieren Sie mögliche Probleme der Validität einer Retestreliabilitätsschätzung.
29. Geben Sie die Definition eines m -Komponententestmodells wieder.
30. Erläutern Sie die Gemeinsamkeiten des Modells multipler Testmessungen und des m -Komponentenmodells.
31. Erläutern Sie die Unterschiede des Modells multipler Testmessungen und des m -Komponentenmodells.
32. Geben Sie die Definition der Reliabilitäten von m -Komponententestmodellen wieder.
33. Skizzieren Sie den intuitiven Zusammenhang zwischen einer Testhalbierungsreliabilität und Cronbach's α .
34. Geben Sie die Definition von Cronbach's α wieder.
35. Erläutern Sie die Intuition zu Cronbach's α basierend auf dem Theorem zur Varianz einer Linearkombination.
36. Geben Sie das Theorem zu Cronbach's α und der Gesamtreliabilität eines m -Komponententestmodells wieder.
37. Wann ist die Gesamtreliabilität eines m -Komponententestmodells mit Cronbach's α identisch?
38. Geben Sie die Definition des Cronbach's α Schätzers wieder.
39. Geben Sie das Theorem zur Spearman-Brown-Formel wieder.
40. Erläutern die Spearman-Brown-Formel mit Bezugnahme auf die Testhalbierungsreliabilität.

Wahrscheinlichkeitstheoretische Ergänzungen

Theorem (Erwartungswert bei linear-affiner Kombination)

ξ sei ein n -dimensionaler Zufallsvektor mit Komponenten $\xi_1, \dots, \xi_n$ und Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \dots \times \mathcal{X}_n$. Weiterhin sei

$$f : \mathcal{X} \rightarrow \mathcal{Z}, x \mapsto f(x) := a_0 + \sum_{i=1}^n a_i x_i \text{ für } a_0, \dots, a_n \in \mathbb{R}. \quad (136)$$

eine linear-affine Kombination der Komponenten von ξ . Dann gilt

$$\mathbb{E}(f(\xi)) = \mathbb{E} \left(a_0 + \sum_{i=1}^n a_i \xi_i \right) = a_0 + \sum_{i=1}^n a_i \mathbb{E}(\xi_i). \quad (137)$$

Beweis

Zur Vereinfachung der Notation schreiben wir $\sum_{x_i \in \mathcal{X}_i}$ als $\sum_{x_i}$.

Beweis (fortgeführt)

$$\begin{aligned}\mathbb{E}(f(\xi)) &= \mathbb{E}\left(a_0 + \sum_{i=1}^n a_i \xi_i\right) \\&= \mathbb{E}(a_0 + a_1 \xi_1 + \cdots + a_n \xi_n) \\&= \sum_{x_1} \cdots \sum_{x_n} (a_0 + a_1 x_1 + \cdots + a_n x_n) p(x_1, \dots, x_n) \\&= \sum_{x_1} \cdots \sum_{x_n} a_0 p(x_1, \dots, x_n) + a_1 x_1 p(x_1, \dots, x_n) + \cdots + a_n x_n p(x_1, \dots, x_n) \\&= \sum_{x_1} \cdots \sum_{x_n} a_0 p(x_1, \dots, x_n) + \sum_{x_1} \cdots \sum_{x_n} a_1 x_1 p(x_1, \dots, x_n) + \cdots + \sum_{x_1} \cdots \sum_{x_n} a_n x_n p(x_1, \dots, x_n) \\&= a_0 \sum_{x_1} \cdots \sum_{x_n} p(x_1, \dots, x_n) + a_1 \sum_{x_1} \cdots \sum_{x_n} x_1 p(x_1, \dots, x_n) + \cdots + a_n \sum_{x_n} \cdots \sum_{x_1} x_n p(x_1, \dots, x_n) \\&= a_0 + a_1 \sum_{x_1} x_1 \sum_{x_2} \cdots \sum_{x_n} p(x_1, \dots, x_n) + \cdots + a_n \sum_{x_n} x_n \sum_{x_{n-1}} \cdots \sum_{x_1} p(x_1, \dots, x_n) \\&= a_0 + a_1 \sum_{x_1} x_1 p(x_1) + \cdots + a_n \sum_{x_n} x_n p(x_n) \\&= a_0 + a_1 \mathbb{E}(\xi_1) + \cdots + a_n \mathbb{E}(\xi_n) \\&= a_0 + \sum_{i=1}^n a_i \mathbb{E}(\xi_i)\end{aligned} \tag{138}$$

Theorem (Iterierter Erwartungswert)

Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2)$. Dann gilt

$$\mathbb{E}(\xi_1) = \mathbb{E}(\mathbb{E}(\xi_1|\xi_2)). \quad (139)$$

Beweis

Wir beschränken uns auf den Beweis für einen diskreten Zufallsvektor. Es gilt

$$\begin{aligned} \mathbb{E}(\xi_1) &= \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1) \\ &= \sum_{x_1 \in \mathcal{X}_1} x_1 \sum_{x_2 \in \mathcal{X}_2} p(x_1, x_2) \\ &= \sum_{x_1 \in \mathcal{X}_1} x_1 \sum_{x_2 \in \mathcal{X}_2} p(x_1|x_2)p(x_2) \\ &= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} x_1 p(x_1|x_2)p(x_2) \\ &= \sum_{x_2 \in \mathcal{X}_2} \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2)p(x_2) \\ &= \sum_{x_2 \in \mathcal{X}_2} \mathbb{E}(\xi_1|\xi_2 = x_2)p(x_2) \\ &= \mathbb{E}(\mathbb{E}(\xi_1|\xi_2)). \end{aligned} \quad (140)$$

Theorem (Nichtnegativität der Varianz)

ξ sei eine Zufallsvariable. Dann gilt

$$\mathbb{V}(\xi) \geq 0. \quad (141)$$

Beweis

Wir betrachten den Fall einer diskreten Zufallsvariable. Dann gilt zunächst

$$(x - \mathbb{E}(\xi))^2 \geq 0 \text{ für alle } x \in \mathcal{X}. \quad (142)$$

Weiterhin gilt für die WMF p von ξ , dass

$$p(x) \geq 0 \text{ für alle } x \in \mathcal{X}. \quad (143)$$

Also folgt

$$p(x)(x - \mathbb{E}(\xi))^2 \geq 0 \text{ für alle } x \in \mathcal{X}. \quad (144)$$

Damit gilt dann aber

$$\mathbb{V}(\xi) = \sum_{x \in \mathcal{X}} p(x)(x - \mathbb{E}(\xi))^2 dx \geq 0, \quad (145)$$

Analog zeigt man die Nichtnegativität der Varianz bei kontinuierlichen Zufallsvariablen.

Theorem (Verschiebungssatz der bedingten Varianz)

Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2)$. Dann gilt

$$\mathbb{V}(\xi_1|\xi_2 = x_2) = \mathbb{E}(\xi_1^2|\xi_2 = x_2) - \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \quad (146)$$

Beweis

Wir beschränken uns auf den Beweis für einen diskreten Zufallsvektor.

$$\begin{aligned} \mathbb{V}(\xi_1|\xi_2 = x_2) &= \mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1|\xi_2 = x_2))^2 | \xi_2 = x_2\right) \\ &= \mathbb{E}\left(\xi_1^2 - 2\xi_1\mathbb{E}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 | \xi_2 = x_2\right) \\ &= \sum_{x_1 \in \mathcal{X}_1} p(x_1|x_2) \left(x_1^2 - 2x_1\mathbb{E}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2\right) \\ &= \sum_{x_1 \in \mathcal{X}_1} x_1^2 p(x_1|x_2) - \sum_{x_1 \in \mathcal{X}_1} 2x_1 \mathbb{E}(\xi_1|\xi_2 = x_2) p(x_1|x_2) + \sum_{x_1 \in \mathcal{X}_1} \mathbb{E}(\xi_1|\xi_2 = x_2)^2 p(x_1|x_2) \\ &= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - 2\mathbb{E}(\xi_1|\xi_2 = x_2) \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \sum_{x_1 \in \mathcal{X}_1} p(x_1|x_2) \\ &= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - 2\mathbb{E}(\xi_1|\xi_2 = x_2)\mathbb{E}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \cdot 1 \\ &= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - 2\mathbb{E}(\xi_1|\xi_2 = x_2)^2 + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \\ &= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \end{aligned}$$

Theorem (Iterierte Varianz)

Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2)$. Dann gilt

$$\mathbb{V}(\xi_1) = \mathbb{E}(\mathbb{V}(\xi_1|\xi_2)) + \mathbb{V}(\mathbb{E}(\xi_1|\xi_2)) \quad (147)$$

Beweis

Wir beschränken uns auf den Beweis für einen diskreten Zufallsvektor. Mit Theorem [Kovarianzverschiebungssatz](#), Theorem [Iterierter Erwartungswert](#) und Theorem [Verschiebungssatz der bedingten Varianz](#) gilt

$$\begin{aligned} \mathbb{V}(\xi_1) &= \mathbb{E}(\xi_1^2) - \mathbb{E}(\xi_1)^2 \\ &= \mathbb{E}(\mathbb{E}(\xi_1^2|\xi_2)) - \mathbb{E}(\mathbb{E}(\xi_1|\xi_2))^2 \\ &= \sum_{x_2 \in \mathcal{X}_2} p(x_2) \mathbb{E}(\xi_1^2|\xi_2 = x_2) - \mathbb{E}(\mathbb{E}(\xi_1|\xi_2))^2 \\ &= \sum_{x_2 \in \mathcal{X}_2} p(x_2) (\mathbb{V}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2) - \mathbb{E}(\mathbb{E}(\xi_1|\xi_2))^2 \\ &= \sum_{x_2 \in \mathcal{X}_2} p(x_2) \mathbb{V}(\xi_1|\xi_2 = x_2) + \sum_{x_2 \in \mathcal{X}_2} p(x_2) \mathbb{E}(\xi_1|\xi_2 = x_2)^2 - \mathbb{E}(\mathbb{E}(\xi_1|\xi_2))^2 \\ &= \mathbb{E}(\mathbb{V}(\xi_1|\xi_2)) + (\mathbb{E}(\mathbb{E}(\xi_1|\xi_2)^2) - \mathbb{E}(\mathbb{E}(\xi_1|\xi_2))^2) \\ &= \mathbb{E}(\mathbb{V}(\xi_1|\xi_2)) + \mathbb{V}(\mathbb{E}(\xi_1|\xi_2)) \end{aligned} \quad (148)$$

Theorem (Kovarianz von Linearkombinationen von Zufallsvariablen)

Gegeben seien n Zufallsvariablen $\xi_1, \dots, \xi_n$ und $n + 1$ reelle Konstanten $a_0, a_1, \dots, a_n$ sowie m Zufallsvariablen $\zeta_1, \dots, \zeta_m$ und $m + 1$ reelle Konstanten $b_0, b_1, \dots, b_m$. Dann gilt

$$\mathbb{C} \left(a_0 + \sum_{i=1}^n a_i \xi_i, b_0 + \sum_{j=1}^m b_j \zeta_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{C}(\xi_i, \zeta_j). \quad (149)$$

Beweis

$$\begin{aligned}
 & \mathbb{C} \left(a_0 + \sum_{i=1}^n a_i \xi_i, b_0 + \sum_{j=1}^m b_j \zeta_j \right) \\
 &= \mathbb{E} \left(\left(a_0 + \sum_{i=1}^n a_i \xi_i - \mathbb{E} \left(a_0 + \sum_{i=1}^n a_i \xi_i \right) \right) \left(b_0 + \sum_{j=1}^m b_j \zeta_j - \mathbb{E} \left(b_0 + \sum_{j=1}^m b_j \zeta_j \right) \right) \right) \\
 &= \mathbb{E} \left(\left(a_0 + \sum_{i=1}^n a_i \xi_i - a_0 - \mathbb{E} \left(\sum_{i=1}^n a_i \xi_i \right) \right) \left(b_0 + \sum_{j=1}^m b_j \zeta_j - b_0 - \mathbb{E} \left(\sum_{j=1}^m b_j \zeta_j \right) \right) \right) \\
 &= \mathbb{E} \left(\left(\sum_{i=1}^n a_i \xi_i - \sum_{i=1}^n a_i \mathbb{E}(\xi_i) \right) \left(\sum_{j=1}^m b_j \zeta_j - \sum_{j=1}^m b_j \mathbb{E}(\zeta_j) \right) \right) \\
 &= \mathbb{E} \left(\sum_{i=1}^n a_i \xi_i \sum_{j=1}^m b_j \zeta_j - \sum_{i=1}^n a_i \xi_i \sum_{j=1}^m b_j \mathbb{E}(\zeta_j) - \sum_{i=1}^n a_i \mathbb{E}(\xi_i) \sum_{j=1}^m b_j \zeta_j + \sum_{i=1}^n a_i \mathbb{E}(\xi_i) \sum_{j=1}^m b_j \mathbb{E}(\zeta_j) \right) \\
 &= \mathbb{E} \left(\sum_{i=1}^n a_i \xi_i \sum_{j=1}^m b_j \zeta_j \right) - \mathbb{E} \left(\sum_{i=1}^n a_i \xi_i \sum_{j=1}^m b_j \mathbb{E}(\zeta_j) \right) - \mathbb{E} \left(\sum_{i=1}^n a_i \mathbb{E}(\xi_i) \sum_{j=1}^m b_j \zeta_j \right) + \mathbb{E} \left(\sum_{i=1}^n a_i \mathbb{E}(\xi_i) \sum_{j=1}^m b_j \mathbb{E}(\zeta_j) \right) \\
 &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{E}(\xi_i \zeta_j) - 2 \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{E}(\xi_i) \mathbb{E}(\zeta_j) + \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{E}(\xi_i) \mathbb{E}(\zeta_j) \\
 &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \left(\mathbb{E}(\xi_i \zeta_j) - 2 \mathbb{E}(\xi_i) \mathbb{E}(\zeta_j) + \mathbb{E}(\xi_i) \mathbb{E}(\zeta_j) \right) \\
 &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{E} \left((\xi_i - \mathbb{E}(\xi_i))(\zeta_j - \mathbb{E}(\zeta_j)) \right) \\
 &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathbb{C}(\xi_i, \zeta_j)
 \end{aligned}$$

Theorem (Kovarianz bei paarweiser Addition)

$\xi_1, \xi_2, \zeta_1, \zeta_2$ seien vier Zufallsvariablen. Dann gilt

$$\mathbb{C}(\xi_1 + \xi_2, \zeta_1 + \zeta_2) = \mathbb{C}(\xi_1, \zeta_1) + \mathbb{C}(\xi_1, \zeta_2) + \mathbb{C}(\xi_2, \zeta_1) + \mathbb{C}(\xi_2, \zeta_2). \quad (150)$$

Beweis

Es seien $n := 2, m := 2, a_0 = b_0 := 0$ und $a_i = b_i := 1$ für $i = 1, 2$. Dann gilt mit Theorem [Kovarianz von Linearkombinationen von Zufallsvariablen](#)

$$\begin{aligned} \mathbb{C}(\xi_1 + \xi_2, \zeta_1 + \zeta_2) &= \mathbb{C}\left(a_0 + \sum_{i=1}^2 a_i \xi_i, b_0 + \sum_{j=1}^2 b_j \zeta_j\right) \\ &= \sum_{i=1}^2 \sum_{j=1}^2 a_i b_j \mathbb{C}(\xi_i, \zeta_j) \\ &= \sum_{i=1}^2 \sum_{j=1}^2 \mathbb{C}(\xi_i, \zeta_j) \\ &= \mathbb{C}(\xi_1, \zeta_1) + \mathbb{C}(\xi_1, \zeta_2) + \mathbb{C}(\xi_2, \zeta_1) + \mathbb{C}(\xi_2, \zeta_2). \end{aligned} \quad (151)$$

Theorem (Symmetrie der Kovarianz)

ξ_1 und ξ_2 seien zwei Zufallsvariablen. Dann gilt

$$\mathbb{C}(\xi_1, \xi_2) = \mathbb{C}(\xi_2, \xi_1) \quad (152)$$

Beweis

Mit der Kommutativität der Multiplikation gilt

$$\mathbb{C}(\xi_1, \xi_2) = \mathbb{E}((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))) = \mathbb{E}((\xi_2 - \mathbb{E}(\xi_2))(\xi_1 - \mathbb{E}(\xi_1))) = \mathbb{C}(\xi_2, \xi_1) \quad (153)$$

□

Theorem (Kovarianzverschiebungssatz)

ξ_1 und ξ_2 seien Zufallsvariablen. Dann gilt

$$\mathbb{C}(\xi_1, \xi_2) = \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2). \quad (154)$$

Beweis

Mit der Definition der Kovarianz gilt

$$\begin{aligned} \mathbb{C}(\xi_1, \xi_2) &= \mathbb{E}((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))) \\ &= \mathbb{E}(\xi_1 \xi_2 - \xi_1 \mathbb{E}(\xi_2) - \mathbb{E}(\xi_1) \xi_2 + \mathbb{E}(\xi_1) \mathbb{E}(\xi_2)) \\ &= \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1) \mathbb{E}(\xi_2) - \mathbb{E}(\xi_1) \mathbb{E}(\xi_2) + \mathbb{E}(\xi_1) \mathbb{E}(\xi_2) \\ &= \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1) \mathbb{E}(\xi_2). \end{aligned} \quad (155)$$

□

Theorem (Varianz einer Linearkombination)

Gegeben seien n Zufallsvariablen $\xi_1, \dots, \xi_n$ und $n + 1$ reelle Konstanten $a_0, a_1, \dots, a_n$. Dann gilt

$$\mathbb{V} \left(a_0 + \sum_{i=1}^n a_i \xi_i \right) = \sum_{i=1}^n a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \mathbb{C}(\xi_i, \xi_j). \quad (156)$$

Beweis

Mit $\mathbb{V}(\xi) = \mathbb{C}(\xi, \xi)$ und Theorem [Kovarianz von Linearkombinationen von Zufallsvariablen](#) gilt zunächst

$$\begin{aligned}\mathbb{V}\left(a_0 + \sum_{i=1}^n a_i \xi_i\right) &= \mathbb{C}\left(a_0 + \sum_{i=1}^n a_i \xi_i, a_0 + \sum_{i=1}^n a_i \xi_i\right) \\&= \mathbb{C}\left(a_0 + \sum_{i=1}^n \xi_i, a_0 + \sum_{j=1}^n a_j \xi_j\right) \\&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^n \sum_{\substack{j=1 \\ j=i}}^n a_i a_j \mathbb{C}(\xi_i, \xi_j) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^n a_i a_i \mathbb{C}(\xi_i, \xi_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^n a_i^2 \mathbb{V}(\xi_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^n a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \mathbb{C}(\xi_i, \xi_j),\end{aligned}\tag{157}$$

Beweis (fortgeführt)

Dabei wurde in der vierten Gleichung die Doppelsumme in solche Terme aufgespalten für die $i = j$ und für die $i \neq j$ und in siebten Gleichung ausgenutzt, dass

$$a_i a_j \mathbb{C}(\xi_i, \xi_j) = a_j a_i \mathbb{C}(\xi_j, \xi_i). \quad (158)$$

Wir verdeutlichen die darausfolgende Identität

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_i a_j \mathbb{C}(\xi_i, \xi_j) = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \mathbb{C}(\xi_i, \xi_j), \quad (159)$$

am Beispiel $n = 3$ untenstehend. Analog mag man sich vorstellen, über alle Elemente außer der Diagonalelemente der $i = 1, \dots, n$ Zeilen und $j = 1, \dots, n$ Spalten einer symmetrischen Matrix mit Einträgen $a_i a_j \mathbb{C}(\xi_i, \xi_j)$ zu summieren. Die linke Seite oberer Gleichung entspricht dann dem zeilenweisen Vorgehen der Summationsbildung für alle Spalteneinträge außer dem, der in der Spalte der jeweils betrachteten Zeile steht, d.h. dem Diagonaleintrag. Die rechte Seite der obigen Gleichung entspricht dann dem Vorgehen, die Symmetrie der Matrix auszunutzen, also die Tatsache zu berücksichtigen, dass die Einträge der Matrix rechts oberhalb und links unterhalb der Diagonalen identisch sind, sodass es genügt, die Einträge der oberen linken Hälfte aufzuaddieren und zu verdoppeln. Dabei werden für jede Zeile $i = 1, \dots, n$ nur gerade die Spalten $j = i + 1, \dots, n$ betrachtet, die rechts von der Diagonale stehen und ihre Einträge aufsummiert. In der letzten Zeile steht dabei nur ein Diagonalelement, das nicht zur Summe gehört, weshalb der Index der äußeren Summe nur bis $n - 1$ läuft.

Beweis (fortgeführt)

Konkret ergibt sich zum Beispiel für $n := 3$:

$$\begin{aligned} & \sum_{i=1}^3 \sum_{\substack{j=1 \\ j \neq i}}^3 a_i a_j \mathbb{C}(\xi_i, \xi_j) \\ &= \sum_{\substack{j=1 \\ j \neq 1}}^3 a_1 a_j \mathbb{C}(\xi_1, \xi_j) + \sum_{\substack{j=1 \\ j \neq 2}}^3 a_2 a_j \mathbb{C}(\xi_2, \xi_j) + \sum_{\substack{j=1 \\ j \neq 3}}^3 a_3 a_j \mathbb{C}(\xi_3, \xi_j) \\ &= a_1 a_2 \mathbb{C}(\xi_1, \xi_2) + a_1 a_3 \mathbb{C}(\xi_1, \xi_3) + a_2 a_1 \mathbb{C}(\xi_2, \xi_1) + a_2 a_3 \mathbb{C}(\xi_2, \xi_3) + a_3 a_1 \mathbb{C}(\xi_3, \xi_1) + a_3 a_2 \mathbb{C}(\xi_3, \xi_2) \\ &= a_1 a_2 \mathbb{C}(\xi_1, \xi_2) + a_2 a_1 \mathbb{C}(\xi_2, \xi_1) + a_1 a_3 \mathbb{C}(\xi_1, \xi_3) + a_3 a_1 \mathbb{C}(\xi_3, \xi_1) + a_2 a_3 \mathbb{C}(\xi_2, \xi_3) + a_3 a_2 \mathbb{C}(\xi_3, \xi_2) \\ &= 2\mathbb{C}(\xi_1, \xi_2) + 2\mathbb{C}(\xi_1, \xi_3) + 2\mathbb{C}(\xi_2, \xi_3) \tag{160} \\ &= 2a_1 a_2 (\mathbb{C}(\xi_1, \xi_2) + a_1 a_3 \mathbb{C}(\xi_1, \xi_3) + a_2 a_3 \mathbb{C}(\xi_2, \xi_3)) \\ &= 2 \left(\sum_{j=2}^3 a_1 a_j \mathbb{C}(\xi_1, \xi_j) + \sum_{j=3}^3 a_2 a_j \mathbb{C}(\xi_2, \xi_j) \right) \\ &= 2 \left(\sum_{j=1+1}^3 a_1 a_j \mathbb{C}(\xi_1, \xi_j) + \sum_{j=2+1}^3 a_2 a_j \mathbb{C}(\xi_2, \xi_j) \right) \\ &= 2 \sum_{i=1}^2 \sum_{j=i+1}^3 a_i a_j \mathbb{C}(\xi_i, \xi_j). \end{aligned}$$

□

Theorem (Varianzen spezieller Linearkombinationen)

(1) (Varianz bei Addition zweier Zufallsvariablen). Gegeben seien zwei Zufallsvariablen ξ und ζ . Dann gilt

$$\mathbb{V}(\xi + \zeta) = \mathbb{V}(\xi) + \mathbb{V}(\zeta) + 2\mathbb{C}(\xi, \zeta). \quad (161)$$

(2) (Varianz bei Subtraktion zweier Zufallsvariablen). Gegeben seien zwei Zufallsvariablen ξ und ζ . Dann gilt

$$\mathbb{V}(\xi - \zeta) = \mathbb{V}(\xi) + \mathbb{V}(\zeta) - 2\mathbb{C}(\xi, \zeta). \quad (162)$$

Beweis

(1) Es seien $n := 2$, $\xi_1 := \xi$, $\xi_2 := \zeta$, $a_0 := 0$, $a_1 := 1$ und $a_2 := 1$. Dann gilt mit Theorem [Varianz einer Linearkombination](#)

$$\begin{aligned}\mathbb{V}(\xi + \zeta) &= \mathbb{V}(a_0 + a_1\xi_1 + a_2\xi_2) \\&= \mathbb{V}\left(a_0 + \sum_{i=1}^2 a_i\xi_i\right) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^{2-1} \sum_{j=i+1}^2 a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^1 \sum_{j=i+1}^2 a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{j=1+1}^2 a_1 a_j \mathbb{C}(\xi_1, \xi_j) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{j=2}^2 a_1 a_j \mathbb{C}(\xi_1, \xi_j) \\&= a_1^2 \mathbb{V}(\xi_1) + a_2^2 \mathbb{V}(\xi_2) + 2a_1 a_2 \mathbb{C}(\xi_1, \xi_2) \\&= 1^2 \cdot \mathbb{V}(\xi) + 1^2 \cdot \mathbb{V}(\zeta) + 2 \cdot 1 \cdot 1 \cdot \mathbb{C}(\xi, \zeta) \\&= \mathbb{V}(\xi) + \mathbb{V}(\zeta) + 2\mathbb{C}(\xi, \zeta).\end{aligned}\tag{163}$$

Beweis (fortgeführt)

(2) Es seien $n := 2$, $\xi_1 := \xi$, $\xi_2 := \zeta$, $a_0 := 0$, $a_1 := 1$ und $a_2 := -1$. Dann gilt mit dem Theorem zur Varianz einer Linearkombination von Zufallsvariablen

$$\begin{aligned}\mathbb{V}(\xi - \zeta) &= \mathbb{V}(a_0 + a_1\xi_1 + a_2\xi_2) \\&= \mathbb{V}\left(a_0 + \sum_{i=1}^2 a_i\xi_i\right) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^{2-1} \sum_{j=i+1}^2 a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^1 \sum_{j=i+1}^2 a_i a_j \mathbb{C}(\xi_i, \xi_j) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{j=1+1}^2 a_i a_j \mathbb{C}(\xi_1, \xi_j) \\&= \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{j=2}^2 a_1 a_j \mathbb{C}(\xi_1, \xi_j) \\&= a_1^2 \mathbb{V}(\xi_1) + a_2^2 \mathbb{V}(\xi_2) + 2a_1 a_2 \mathbb{C}(\xi_1, \xi_2) \\&= 1^2 \cdot \mathbb{V}(\xi) + (-1)^2 \cdot \mathbb{V}(\zeta) + 2 \cdot 1 \cdot (-1) \cdot \mathbb{C}(\xi, \zeta) \\&= \mathbb{V}(\xi) + \mathbb{V}(\zeta) - 2\mathbb{C}(\xi, \zeta).\end{aligned}\tag{164}$$

□

Theorem (Verschiebungssatz der bedingten Kovarianz)

Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2, \xi_3)$. Dann gilt

$$\mathbb{C}(\xi_1, \xi_2 | \xi_3) = \mathbb{E}(\xi_1 \xi_2 | \xi_3) - \mathbb{E}(\xi_1 | \xi_3) \mathbb{E}(\xi_2 | \xi_3). \quad (165)$$

Beweis

Ein Beweis ergibt sich durch die Ersetzung der entsprechenden Erwartungswerte im Beweis von Theorem [Kovarianzverschiebungssatz](#) durch die entsprechenden bedingten Erwartungswerte.

□

- Borsboom, D. (2009). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 1904-1920, 3(3), 296–322. <https://doi.org/10.1111/j.2044-8295.1910.tb00207.x>
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Feldt, L. S. (1965). The Approximate Sampling Distribution of Kuder-Richardson Reliability Coefficient Twenty. *Psychometrika*, 30(3), 357–370. <https://doi.org/10.1007/BF02289499>
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement*, 11(1), 93–103. <https://doi.org/10.1177/014662168701100107>
- Foster, R. C. (2020). A generalized framework for classical test theory. *Journal of Mathematical Psychology*, 96, 102330. <https://doi.org/10.1016/j.jmp.2020.102330>
- Gulliksen, H. (1950). *Theory of mental tests*. John Wiley & Sons Inc. <https://doi.org/10.1037/13240-000>
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255–282. <https://doi.org/10.1007/BF02288892>
- Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI-II Beck Depressions-Inventar*. Pearson.
- Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6(3), 153–160.
- Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, 28(3), 221–2238. <https://doi.org/10.1007/BF02289571>

- Lazarsfeld, P. (1959). Latent structure analysis. In *Psychology: A study of a science*, Vol. 3. McGraw-Hill.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores* (Nachdr. der Ausg. Reading, Mass. [u.a.], 1968). Information Age Publ.
- Moosbrugger, H., & Kelava, A. (Eds.). (2012). *Testtheorie und Fragebogenkonstruktion: mit 66 Abbildungen und 41 Tabellen* (2., aktualisierte und überarbeitete Auflage). Springer.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32(1), 1–13.
- Pospeschill, M. (2022). *Testtheorie, Testkonstruktion, Testevaluation* (2nd ed.). Ernst Reinhardt Verlag.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 1904-1920, 3(3), 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the Distribution of the Maximum Likelihood Estimator of Cronbach's Alpha. *Psychometrika*, 65(3), 271–280. <https://doi.org/10.1007/BF02296146>
- Yuan, K.-H., & Bentler, P. M. (2002). On Robustness of the Normal-Theory Based Asymptotic Distributions of Three Reliability Coefficient Estimates. *Psychometrika*, 67(2), 251–259. <https://doi.org/10.1007/BF02294845>