Testtheorie und Testkonstruktion

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(3) Modellformulierung

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 $\text{r}_{\text{tt}} = .93$. Aus Studentenstichproben werden für ein- bis zweiwöchige Zeiträume Koeffizienten von $\text{r}_{\text{tt}} = .74 - .96$ 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, Keller, und Kühner (2006)

Klassische Testtheorie

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

$$\text{Observed Score} = \text{True Score} + \text{Error Score} \tag{1}$$

- Modellierung von intra- und interindividueller Variabilität
- Grundlage für die quantitative Reliabiliätsbeurteilung

$\Rightarrow$ Paralleltestreliabiliät, Spearman-Brown-Formel, Cronbach's α

Die Klassische Testtheorie ist zunächst einmal eine Reliabilitätstheorie

Wir gehen davon aus, dass Begrifflichkeiten wie *Paralleltestreliabilität, Retestreliabilität, Cronbach's* α*, Spearman-Brown-Formel* aus dem Bachelor schon bekannt, aber nicht wirklich formal verstanden sind.

Insbesondere wollen wir im Seminar folgende Partikularitäten der Klassischen Testheorie erläutern:

- Warum kann die Reliabilität eines Tests einerseits über die Korrelation von True-Scores und Observed Scores definiert sein, gleichzeitig aber allein über die Korrelation der Observed Scores mehrerer Tests bestimmt werden?
- Woher kommt die Formel für Cronbach's α und wann und warum gilt sie?
- Woher kommt die Spearman-Brown Formel und wann und warum gilt sie?

Zur Beantwortung dieser Fragen müssen wir einiges an Vorarbeit leisten

(1) Wiederholung allgemeiner Grundlagen zu Erwartungswerten, Varianzen, Kovarianzen, Korrelation
(2) Explizite Formulierung des Modells der Klassischen Testtheorie
(3) Definition von Paralleltests
(4) Definition der Reliabilität

Basierend auf diesen Vorarbeiten sind dann Antworten auf obige Fragen möglich.

Modellformulierung

Erste Modelleigenschaften

Selbstkontrollfragen

Modellformulierung

Erste Modelleigenschaften

Selbstkontrollfragen

Definition (True-Score, Observed-Score und Error-Score)

Für $i = 1, ..., n$ und $j = 1, ..., m$ ist der *True-Score* t_{ij} einer Person i für eine Testmessung j definiert als der bedingte Erwartungswert des *Observed Scores* der Person für diese Testmessung

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

Der *Error-Score* einer Person ist definiert als die Zufallsvariable

$$\varepsilon_{ij} := \upsilon_{ij} - t_{ij}. \tag{3}$$

Bemerkungen

- True-, Observed- und Error-Score werden auch *wahrer Wert*, *beobachteter Wert* und *Messfehler* genannt.
- Eine *Testmessung* mag aus einem Item oder der Summe mehrer Items bestehen.
- Man beachte, dass τ_{ij}, υ_{ij} und ε_{ij} hier Zufallsvariablen sind und $t_{ij} \in \mathbb{R}$ eine Konstante ist.
- Die Definition von t_{ij} ist zirkulär bis tautologisch, da t_{ij} mithilfe von t_{ij} definiert wird.
- Die bedingte Verteilung $\mathbb{P}(\upsilon_{ij}|\tau_{ij} = t_{ij})$ wird *Propensitätsverteilung* genannt.
- Die Propensitätsverteilung modelliert intraindividuelle Observed-Score Variabilität bei festem True-Score.

Modellformulierung

Zur Interpretation von Propensitätsverteilung und True Score

"Suppose we ask an individual, Mr. Brown, repeatedly whether he is in favor of the United Nations; suppose further that after each question we"wash his brains" and ask him the same question again. Because Mr. Brown is not certain as to how he feels about the United Nations, he will sometimes give a favorable and sometimes an unfavorable answer. Having gone through this procedure many times, we then compute the proportion of times Mr. Brown was in favor of the United Nations."

(Lazarsfeld, 1959; zitiert in Lord und Novick (1968), pp.29 - 30)

Aber

… Propensität ist eigentlich eine objektivistische Interpretation von Wahrscheinlichkeiten als "Verwirklichungstendenz"
… und im Gegensatz zu streng Frequentistischen Theorien ergeben Propensitäten auch im Einzelfall Sinn.

Für eine ausführliche Diskussion zur Interpretation Modellwahrscheinlichkeiten, siehe Borsboom (2009).

Modellformulierung

Definition (Modell multipler Testmessungen)

Für $i = 1, ..., n$ und $j = 1, ..., m$ sei

- τ_{ij} eine Zufallsvariable, die den *True-Score* der $i$ten Person in der $j$ten Testmessung modelliere,
- υ_{ij} eine Zufallsvariable, die den *Observed-Score* der $i$ten Person in der $j$ten Testmessung modelliere,

Dann nennen wir die gemeinsame Verteilung der τ_{ij} und υ_{ij} mit der Faktorisierungseigenschaft

$$\mathbb{P}\left(\tau_{11}, \upsilon_{11}, ..., \tau_{nm}, \upsilon_{nm}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_{i1}, ..., \tau_{im}) \prod_{j=1}^{m} \mathbb{P}(\upsilon_{ij}|\tau_{ij}) \tag{4}$$

das *Modell multipler Testmessungen*, wenn gilt, dass

$$\mathbb{P}(\tau_{11}, ..., \tau_{1m}) \prod_{j=1}^{m} \mathbb{P}(\upsilon_{1j}|\tau_{1j}) = \cdots = \mathbb{P}(\tau_{n1}, ..., \tau_{nm}) \prod_{j=1}^{m} \mathbb{P}(\upsilon_{nj}|\tau_{nj}). \tag{5}$$

Bemerkungen

- Die gemeinsame Verteilung der τ_{ij}, υ_{ij} faktorisiert über $i = 1, ..., n$
- $\Leftrightarrow$ Unabhängkeitsannahme über Personen hinsichtlich der True-Scores und Observed-Scores
- Für $i = 1, ..., n$ faktorisiert die gemeinsame Verteilung der $\tau_{i1}, ..., \tau_{im}$ über $j = 1, ..., m$ nicht immer.
- $\Leftrightarrow$ Die True-Scores einer Person können zusammenhängen (Parallelität, τ-Äquivalenz, etc.)

Modellformulierung

Bemerkungen (fortgesetzt)

- Für jedes $i = 1, ..., n$ sind die υ_{ij} für $j = 1, ..., m$ bedingt unabhängig gegeben τ_{ij}.
- $\Leftrightarrow$ Der True-Score in Testmessung $k \neq j$ beinflusst den Observed-Score in Testmessung j nicht.
- $\Leftrightarrow$ Der Observed-Score in Testmessung $k \neq j$ beinflusst den Observed-Score in Testmessung j nicht.
- Die Marginalverteilungen $\mathbb{P}(\tau_{i1}, ..., \tau_{im}) \prod_{j=1}^{m} \mathbb{P}(\upsilon_{ij}|\tau_{ij})$ sind identisch für $i = 1, ..., n$
- $\Leftrightarrow$ Die Daten einer Person werden modelliert als unabhängige Realisierungen aus einer "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1}, ..., \tau_{\bullet m}) \prod_{j=1}^{m} \mathbb{P}(\upsilon_{\bullet j}|\tau_{\bullet j}) \tag{6}$$

- Für eine Testmessung gilt

$$\mathbb{P}\left(\tau_{11}, \upsilon_{11}, ..., \tau_{n1}, \upsilon_{n1}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_{i1})\mathbb{P}(\upsilon_{i1}|\tau_{i1}) \tag{7}$$

- $\Leftrightarrow$ Die Daten einer Person für eine Testmessung sind unabhängige Realisierungen der "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1})\mathbb{P}(\upsilon_{\bullet 1}|\tau_{\bullet 1}) \tag{8}$$

- $\Leftrightarrow$ Die i, jte Marginalverteilung des Modells multipler Testmessungen hat die Form (bitte beweisen!)

$$\mathbb{P}(\tau_{ij})\mathbb{P}(\upsilon_{ij}|\tau_{ij}) \tag{9}$$

Generative Sichtweise des Modells multipler Testmessungen für eine Testmessung (Lord and Novick (1968), S. 34)

- Für Person i und Testmessung j wird zunächst ein True-Score t_{ij} anhand von $\mathbb{P}(\tau_{ij})$ realisiert.
- Für Person i und Testmessung j wird dann ein Observed-Score υ_{ij} anhand von $\mathbb{P}(\upsilon_{ij}|\tau_{ij} = t_{ij})$ realisiert.
- Genauer wird eigentlich eine Testmessung j festgelegt, ein Personenindex i realisiert und t_{ij} evaluiert.

Definition (Einfaches Modell der Klassischen Testtheorie)

τ sei eine Zufallsvariable, die die Verteilung der *True-Scores* zu einer Testmessung in einer Population beschreibe und y sei eine Zufallsvariable, die die Verteilung der *Observed-Scores* einer Testmessung beschreibe. Dann heißt die gemeinsame Verteilung von τ und y,

$$\mathbb{P}(\tau, \upsilon) = \mathbb{P}(\tau)\mathbb{P}(\upsilon|\tau) \tag{10}$$

das *Einfache Modell der Klassischen Testtheorie* für eine Testmessung.

Bemerkungen

- Gegenüber dem Modell multipler Testmessungen treten weniger Zufallsvariablen und Indizes auf.
- Die Redundanz der Gleichheit vieler Verteilungen im Modell multipler Testmesssung tritt nicht auf.
- n beobachtete Testwerte werden als Teilbeobachtungen von Ziehungen aus $\mathbb{P}(\tau, \upsilon)$ gedacht.
- Für die Unabhängigkeit über Personen mag man auch $(\tau_1, \upsilon_1), ..., (\tau_n, \upsilon_n) \sim \mathbb{P}(\tau, \upsilon)$ schreiben.
- Viele wichtige Eigenschaften des Modells der Klassischen Testtheorie ergeben sich schon basierend auf $\mathbb{P}(\tau, \upsilon)$.
- Das Modell sieht einfacher zu handhaben aus als das Modell multipler Testmessungen
- Man findet das Einfache Modell der Klassischen Testtheorie oft in Lehrbüchern.
- Aber: Wenn mehrere Personen und mehrere Tests in Spiel kommen wird unklar, wovon die Rede ist.
- Aber: Schätzer der Modellparameter beruhen immer auf allen Observed-Score Zufallsvariablen $\upsilon_{1j}, ...\upsilon_{nj}$.
- Wir werden nach Bedarf zwischen beiden Modellformulierungen hin und her wechseln.
- Die Modellformulierung in Sinne des Modells multipler Testmessungen ist unser Standardfall.

Modellformulierung

Erste Modelleigenschaften

Selbstkontrollfragen

Erste Modelleigenschaften

Theorem (Erste Eigenschaften des Modells multipler Testmessungen)

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

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

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

Bemerkungen

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

$$\mathbb{E}(\varepsilon|\tau = t) = 0, \quad \mathbb{E}(\varepsilon) = 0, \quad \mathbb{C}(\tau, \varepsilon) = 0, \quad \mathbb{V}(\upsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) \text{ und } \mathbb{C}(\upsilon, \tau) = \mathbb{V}(\tau) \tag{11}$$

geschrieben.

Erste Modelleigenschaften

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

- Die Tatsache, dass der bedingte Erwartungswert des Error-Scores 0 ist, folgt aus der True-Score Definition.
- Dies ist ein anderer Ansatz als die Annahme, dass der bedingte Erwartungswert des Error-Scores 0 ist.
- Weil der bedingte Error-Score Erwartungswert für jeden True-Score 0 ist, ist auch sein Erwartungswert 0.

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

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

Bemerkungen zu $\mathbb{V}(\upsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon)$

- Die Observed-Score Varianz kann additiv in Beiträge der True-Score und der Error-Score Varianz zerlegt werden.
- Diese Eigenschaft des Modells ist für eine Eigenschaft der Reliabilität essentiell.

Bemerkungen zu $\mathbb{C}(\upsilon, \tau) = \mathbb{V}(\tau)$

- Die Kovarianz der Observed Scores und True Scores einer Testmessung entspricht der Varianz der True Scores.
- Diese Eigenschaft des Modells ist für eine weitere Eigenschaft der Reliabilität essentiell.

Beweis

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

$$\upsilon := \upsilon_{ij}, \tau := \tau_{ij} \text{ und analog } \upsilon := \upsilon_{ij}, t := t_{ij} \text{ sowie } \mathcal{Y} := \mathcal{Y}_{ij}, \mathcal{T} := \mathcal{T}_{ij} \tag{12}$$

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

$$p(t, y) = p(y|t)p(t) \tag{13}$$

voraus. Der kontinuierliche Fall folgt dann jeweils analog.

(1) Es gilt

$$\begin{aligned}
\mathbb{E}(\varepsilon|\tau = t) &:= \mathbb{E}(\upsilon - \tau|\tau = t) \\
&= \sum_{y\in\mathcal{Y}} (y - t)\, p(y|t) \\
&= \sum_{y\in\mathcal{Y}} yp(y|t) - \sum_{y\in\mathcal{Y}} tp(y|t) \\
&= \mathbb{E}(\upsilon|\tau = t) - t \sum_{y\in\mathcal{Y}} p(y|t) \\
&= t - t \cdot 1 \\
&= 0.
\end{aligned} \tag{14}$$

Erste Modelleigenschaften

Beweis (fortgeführt)

(2) Es gilt

$$
\begin{aligned}
\mathbb{E}(\varepsilon) &:= \mathbb{E}(\upsilon - \tau) \\
&= \sum_{y \in \mathcal{Y}} \sum_{t \in \mathcal{T}} (y - t)\, p(t, y) \\
&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} (y - t)\, p(y|t) p(t) \\
&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} y p(y|t) p(t) - t p(y|t) p(t) \\
&= \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{Y}} p(t) \left(y p(y|t) - t p(y|t) \right) \\
&= \sum_{t \in \mathcal{T}} p(t) \left(\sum_{y \in \mathcal{Y}} y p(y|t) - t \sum_{y \in \mathcal{Y}} p(y|t) \right) \\
&= \sum_{t \in \mathcal{T}} p(t) \left(t - t \cdot 1 \right) \\
&= \sum_{t \in \mathcal{T}} p(t) \cdot 0 \\
&= 0.
\end{aligned}
\tag{15}
$$

Beweis (fortgeführt)

(3) Es gilt

$$
\begin{aligned}
\mathbb{C}(\tau,\varepsilon) & = \mathbb{E}\left((\tau-\mathbb{E}(\tau))(\varepsilon-\mathbb{E}(\varepsilon))\right) \\
& = \mathbb{E}\left((\tau-\mathbb{E}(\tau))\varepsilon\right) \\
& = \mathbb{E}\left((\tau-\mathbb{E}(\tau))(\upsilon-\tau)\right) \\
& = \sum_{y\in\mathcal{Y}}\sum_{t\in\mathcal{T}}\left((t-\mathbb{E}(\tau))(y-t)\right)p(t,y) \\
& = \sum_{t\in\mathcal{T}}\sum_{y\in\mathcal{Y}}(t-\mathbb{E}(\tau))(y-t)p(y|t)p(t) \\
& = \sum_{t\in\mathcal{T}}p(t)(t-\mathbb{E}(\tau))\sum_{y\in\mathcal{Y}}(y-t)\,p(y|t) \\
& = \sum_{t\in\mathcal{T}}p(t)(t-\mathbb{E}(\tau))\sum_{y\in\mathcal{Y}}(yp(y|t)-tp(y|t)) \\
& = \sum_{t\in\mathcal{T}}p(t)(t-\mathbb{E}(\tau))\left(\sum_{y\in\mathcal{Y}}yp(y|t)-t\sum_{y\in\mathcal{Y}}p(y|t)\right) \\
& = \sum_{t\in\mathcal{T}}p(t)(t-\mathbb{E}(\tau))\,(t-t\cdot 1) \\
& = \sum_{t\in\mathcal{T}}p(t)(t-\mathbb{E}(\tau))\cdot 0 \\
& = 0.
\end{aligned}
\tag{16}
$$

Erste Modelleigenschaften

Beweis (fortgeführt)

(4) Mit dem Theorem zu Varianzen von Summen und Differenzen von Zufallsvariablen sowie Aussage (3) des Theorems gilt

$$\mathbb{V}(\upsilon) = \mathbb{V}(\tau + \varepsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) + 2\mathbb{C}(\tau, \varepsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) + 2 \cdot 0 = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) \tag{17}$$

(5) Mit dem Kovarianzverschiebungssatz,der Linearkombinationseigenschaft des Erwartungswerts, Aussage (2) des Theorems und der Tatsache, dass mit Aussage (4) des Theorems außerdem folgt, dass

$$\mathbb{E}(\varepsilon\tau) = \mathbb{E}(\tau\varepsilon) = \mathbb{C}(\tau, \varepsilon) + \mathbb{E}(\tau)\mathbb{E}(\varepsilon) = 0 + \mathbb{E}(\tau) \cdot 0 = 0 \tag{18}$$

gilt

$$\begin{aligned}
\mathbb{C}(\upsilon, \tau) &= \mathbb{E}(\upsilon\tau) - \mathbb{E}(\upsilon)\mathbb{E}(\tau) \\
&= \mathbb{E}((\tau + \varepsilon)\tau) - \mathbb{E}(\tau + \varepsilon)\mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2 + \varepsilon\tau) - (\mathbb{E}(\tau) + \mathbb{E}(\varepsilon))\,\mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) + \mathbb{E}(\varepsilon\tau) - \mathbb{E}(\tau)^2 - \mathbb{E}(\varepsilon)\mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) + \mathbb{E}(\varepsilon\tau) - \mathbb{E}(\tau)^2 - 0 \cdot \mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) + 0 - \mathbb{E}(\tau)^2 - 0 \cdot \mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) - \mathbb{E}(\tau)^2 \\
&= \mathbb{V}(\tau).
\end{aligned} \tag{19}$$

Erste Modelleigenschaften

Beispiel (Lord und Novick (1968), Exercise 2.17)

Es sei $i = 1, ..., n$, $m := 1$ mit

$$\mathbb{P}(\tau_{i1}) := N(\mu, 1) \text{ und } \mathbb{P}(\upsilon_{i1}|\tau_{i1}) := N(\tau_{i1}, 1) \tag{20}$$

Dann gelten

(1) $\mathbb{P}(\upsilon_{i1}) = N(\mu, 2)$
(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

- Die Beispieldefinitionen induzieren eine gemeinsame Normalverteilungen von τ_{i1} und υ_{i1}.
- Die Beispieldefinitionen und $\varepsilon_{i1} := \upsilon_{i1} - \mathbb{E}(\upsilon_{i1}|\tau_{i1})$ 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.

Beweis

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

(1) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(\upsilon|\tau) = N(\tau, 1) \tag{21}$$

induzierte gemeinsame Verteilung von τ und y, wobei offenbar

$$\mathbb{P}(\upsilon|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 1 \text{ und } b := 0 \tag{22}$$

gilt. Aus dem Theorem zu gemeinsamen Normalverteilungen ergibt sich dann zunächst, dass

$$\begin{pmatrix} \tau \\ \upsilon \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ 1 \cdot \mu + 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \cdot 1 \\ 1 \cdot 1 & 1 + 1 \cdot 1 \cdot 1 \end{pmatrix}\right) = N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right) \tag{23}$$

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

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

$$\varepsilon := \upsilon - t \text{ mit } \upsilon \sim N(t, 1) \tag{24}$$

Mit dem Theorem zu linear-affinen Transformation einer normalverteilten Zufallsvariable gilt dann

$$\varepsilon \sim N\left(t - t, 1^2 \cdot 1\right) = N(0, 1) \tag{25}$$

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

Erste Modelleigenschaften

Beweis (fortgesetzt)

(3) und (4) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(\varepsilon|\tau) = N(0, 1) \tag{26}$$

induzierte gemeinsame Verteilung von τ und ε, wobei offenbar

$$\mathbb{P}(\varepsilon|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 0 \text{ und } b := 0 \tag{27}$$

gilt. Aus dem Theorem zu gemeinsamen Normalverteilungen ergibt sich dann zunächst, dass

$$\begin{pmatrix} \tau \\ \varepsilon \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ 0 \cdot \mu + 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \cdot 0 \\ 0 \cdot 1 & 1 + 0 \cdot 1 \cdot 0 \end{pmatrix}\right) = N\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \tag{28}$$

Aus dem Theorem zu marginalen Normalverteilungen ergeben sich dann durch Ablesen

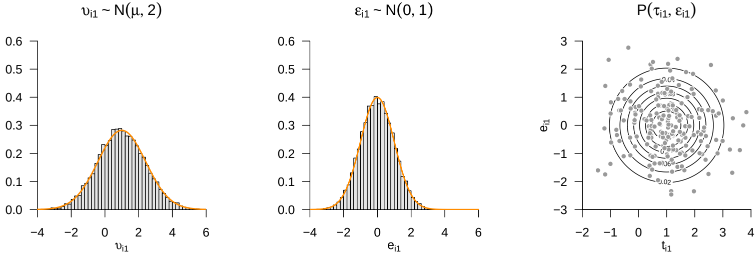
$$\varepsilon \sim N(0, 1) \text{ und } \mathbb{C}(\tau, \varepsilon) = 0. \tag{29}$$

Erste Modelleigenschaften

Simulation mit $\mu := 1, n := 10^4$

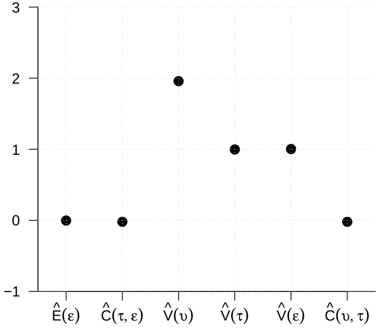
```
n           = 1e4                            # Personenanzahl
m           = 1                              # Testmessungsanzahl
mu          = 1                              # True-Score Erwartungswertparameter
T           = matrix(rep(NaN, n*m), nrow = n) # True-Score Array
Y           = matrix(rep(NaN, n*m), nrow = n) # Observed-Score Array
E           = matrix(rep(NaN, n*m), nrow = n) # Error-Score Array
for(i in 1:n){                               # Iteration über Personen
  for(j in 1:m){                             # Iteration über Testmessungen
    T[i,j]  = rnorm(1,mu,1)                  # True-Score Realisierung
    Y[i,j]  = rnorm(1,T[i,j],1)              # Observed-Score Realisierung
    E[i,j]  = Y[i,j] - T[i,j]}}              # Error-Score Realisierung
e_hat_es    = mean(E[,1])                    # Erwartungswertschätzung Error-Score
c_hat_ts_es = cov(T[,1],E[,1])               # Kovarianzschätzung True Score, Error-Score
v_hat_os    = var(Y[,1])                     # Varianzschätzung Observed-Score
v_hat_ts    = var(T[,1])                     # Varianzschätzung True-Score
v_hat_es    = var(E[,1])                     # Varianzschätzung Error-Score
c_hat_os_ts = cov(T[,1],E[,1])               # Kovarianzschätzung Observed-Score, True-Score
```

Simulation mit $\mu := 1, n := 10^4$



Erste Modelleigenschaften

Simulation mit $\mu := 1, n := 10^4$



Modellformulierung

Erste Modelleigenschaften

Selbstkontrollfragen

Selbskontrollfragen

1. Geben Sie die Definition von True-Score, Observed-Score und Error-Score wieder.
2. Geben Sie die Definition des Einfachen Modells der Klassischen Testtheorie wieder.
3. Geben Sie das Theorem zu ersten Eigenschaften des Modells multipler Testmessungen wieder.

Referenzen

Borsboom, Denny. 2009. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.

Gulliksen, Harold. 1950. *Theory of Mental Tests*. Hoboken: John Wiley & Sons Inc. https://doi.org/10.1037/13240-000.

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

Lord, Frederic M., und Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Nachdr. der Ausg. Reading, Mass. [u.a.], 1968. The Addison-Wesley Series in Behavioral Science: Quantitative Methods. Charlotte, NC: Information Age Publ.