Probabilistische Datenwissenschaft für die Psychologie

Dirk Ostwald

Inhaltsverzeichnis

Willkommen

Herzlich willkommen zur Arbeitsversion von *Probabilistische Datenwissenschaft für die Psychologie (PDWP)*, einem Lehrbuch zur datenwissenschaftlichen Methodenlehre am Institut für Psychologie der Otto-von-Guericke-Universität Magdeburg.

Zitation

Ostwald, D. (2024) Probabilistische Datenwissenschaft für die Psychologie. 10.5281/zenodo.10730199

Lizenz

Teil I.

1. Klassische Testtheorie

–>

Notation

Zufallsvariablen y, τ, ε mit Werten y, t, e in Y, T, E

1.1. Das Modell multipler Testmessungen

Die Klassische Testtheorie in der Formalisierung nach Novick (1966) und Lord & Novick (1968) nimmt ihren Ausgang von der Definition des *True-Scores* einer Testmessung einer Person, die die Definitionen des *Observed-Scores* und des *Error-Scores* impliziert. Sie hat folgende Form.

Definition 1.1 (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

$$t_{ij} := \mathbb{E}(y_{ij} | \tau_{ij} = t_{ij}). \tag{1.1}$$

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

$$\varepsilon_{ij} := y_{ij} - t_{ij}. \tag{1.2}$$

•

True-, Observed- und Error-Score werden im Deutschen natürlich auch *wahrer Wert*, *beobachteter Wert* und *Messfehler* genannt. Mit dem Begriff einer *Testmessung* ist hier etwas unspezifisch und je nach Anwendung entweder eine Messung mithilfe eines einzelnen Items oder mithilfe der Summe mehrer Items gemeint. An späterer Stelle werden wir die Unterscheidung von Itemscores und Testsummenscores explizit machen, welche im Rahmen der Klassischen Testtheorie den Begriff der m-Komponententestmodelle (vgl. Kapitel 1.4) induziert. Man beachte, dass der Definition bedingter Erwartungswerte gemäß τ_{ij}, y_{ij} und ε_{ij} hier Zufallsvariablen sind und $t_{ij} \in \mathbb{R}$ eine Konstante ist.

Die Definition des True-Scores t_{ij} in Definition 1.1 ist etwas speziell, um nicht zu sagen zirkulär bis tautologisch, da t_{ij} mithilfe von t_{ij} definiert wird. Die zentrale Motivation von Novick (1966) und Lord & Novick (1968) den True-Scores t_{ij} als bedingten Erwartungswert, anstelle von zum Beispiel einfach einer Realisierung der Zufallvariable tau_{ij} zu definieren, war es, sich einer Diskussion zur metaphysischen Bedeutung eines True-Scores zu entziehen. Hätten Novick (1966) und Lord & Novick (1968) beispielsweise den

True-Score als Realisierung einer latenten Variable definiert, die einer Person für eine bestimmte Testmessung eigen sein soll, so hätte dies in der zeitgenössischen Diskussion eindeutig den Charakter einer angreifbaren metaphysischen Aussage. Stattdessen versuchen Novick (1966) und Lord & Novick (1968) in ihrer Definition möglichst operationalistisch vorzugehen und den True Score einer Person für eine Testmessung als “Durchschnitt der Observed Scores” einer Person über “wiederholte Testmessungen unter identischen Bedingungen” darzustellen. Zur Bedeutung des True-Scores zitieren Lord & Novick (1968), S.29 - 30 Lazarsfeld, 1959

“Angenommen, wir fragen eine Person, 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 True-Score verstehen (vgl. auch Borsboom et al. (2004)). Natürlich ist dieser Versuch einer antimetaphysischen Grundhaltung nicht durchhaltbar: zum einen handelt es sich bei der “wiederholten Testmessung unter identischen Bedingugen” um ein idealisiertes, in der Realität nicht durchführbares Gedankenexperiment. Zum anderen definiern Novick (1966) und Lord & Novick (1968) den True-Score auch gerade nicht als (endlichen) Mittelwert einer Messreihe, sondern als idealisierten Erwartungswert einer Zufallsvariable. Als wirklich operationalistische, Metaphysik-freie Definition überzeugt Definition 1.1 also nicht, verkompliziert die Entwicklung einer Standardmessfehlertheorie, wie sie beispielsweise dies klassische Formalisierung des ALMs darstellt, für die Analyse von Tests und Fragebögen aber beträchtlich. Dies mag einer der Gründe sein, warum die Klassische Testtheorie bis heute lediglich in der Psychologie und wenig darüberhinaus von Bedeutung ist.

Ähnlich gelagert ist die Bezeichnung der bedingten Verteilung des Observed-Scores $\mathbb{P}(y_{ij}|\tau_{ij} = t_{ij})$ als *Propensitätsverteilung* durch Lord & Novick (1968), welche die intraindividuelle Observed-Score Variabilität bei festem True-Score modelliert. Dabei klingt an, dass Lord & Novick (1968) eine Propensitätsinterpretation von Wahrscheinlichkeiten als kausal bedingte “Verwicklungstendenzen”, die, im Gegensatz zur Frequentistischen Interpretation auch im Einzelfall Sinn ergibt. 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)).

In unserer Darstellung wollen wir dem zeitgemäßen modell-basiert-realistischem Ansatz folgen und wählen mit Definition 1.2 deshalb eine Formulierung des Modells multipler Testmessungen der Klassischen Testtheorie, die mit Definition 1.1 und damit natürlich auch den theoretischen Ergebnissen der Klassischen Testtheorie kongruent ist, aber nicht versucht, ihren Modellcharakter zu verschleiern und dabei insbesondere auch das Gesamtziel der Modellierung einer Menge von m Testmessungen von n Personen, als eines Datensatzes von nm Datenpunkten in den Vordergrund stellt. Wir definieren das *Modell multipler Testmessungen* daher wie folgt.

Definition 1.2 (Modell multipler Testmessungen)**.** Für $i = 1, ..., n$ und $j = 1, ..., m$ seien τ_{ij} eine Zufallsvariable, die den *True-Score* der $i$ten Person in der $j$ten Testmessung modelliere und y_{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 y_{ij} mit der Faktorisierungseigenschaft

$$\mathbb{P}\left(\tau_{11}, y_{11}, ..., \tau_{nm}, y_{nm}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_{i1}, ..., \tau_{im}) \prod_{j=1}^{m} \mathbb{P}(y_{ij}|\tau_{ij}) \tag{1.3}$$

das *Modell multipler Testmessungen*, wenn gilt, dass

$$\mathbb{P}(\tau_{11}, ..., \tau_{1m}) \prod_{j=1}^{m} \mathbb{P}(y_{1j}|\tau_{1j}) = \cdots = \mathbb{P}(\tau_{n1}, ..., \tau_{nm}) \prod_{j=1}^{m} \mathbb{P}(y_{nj}|\tau_{nj}). \tag{1.4}$$

•

Die Definition des Modells multipler Testmessungen bildet einige grundlegende Annahmen zur Unabhängigkeit und Identität von Verteilungen in der Klassischen Testtheorie ab. Zunächst einmal wird angenommen, dass die gemeinsame Verteilung der τ_{ij}, y_{ij} über $i = 1, ..., n$ faktorisiert, dass die Verteilungen der True-Scores und Observed-Scores also über Personen unabhängig sind. Wissen um True- oder Observed-Scores einer Person ändert die angenommenen Verteilungen der True- oder Observed-Scores anderer Personen also nicht. Dagegen faktorisiert für jedes $i = 1, ..., n$ die gemeinsame Verteilung der $\tau_{i1}, ..., \tau_{im}$ über Testmessungen $j = 1, ..., m$ nicht notwendigerweise. Die True-Scores einer Person können also abhängig sein und damit Wissen um die Ausprägung einer Testmessung bei einer Person die Verteilung des True-Scores anderen Testmessung derselben Person informieren. In der Klassischen Testtheorie werden verschiedene Arten dieser Form von Abhängigkeiten unterschieden und wie wir später sehen werden beispielsweise als *Parallelität*, *τ-Äquivalenz* und *essentielle τ-Äquivalenz* bezeichnet. Weiterhin wird für jede Person $i = 1, ..., n$ angenommen, dass die Observed-Scores y_{ij} für $j = 1, ..., m$ gegeben τ_{ij} bedingt unabhängig sind. Dies impliziert einerseits, dass für eine Person der True-Score in Testmessung $k \neq j$ den Observed-Score in Testmessung j nicht beinflusst und ebenso, dass der Observed-Score in Testmessung $k \neq j$ den Observed-Score in Testmessung j nicht beeinflusst.

Schließlich wird angenommen, dass die Marginalverteilungen

$$\mathbb{P}(\tau_{i1}, y_{i1}, ..., \tau_{im}, y_{im}) = \mathbb{P}(\tau_{i1}, ..., \tau_{im}) \prod_{j=1}^{m} \mathbb{P}(y_{ij}|\tau_{ij}) \tag{1.5}$$

über Personen $i = 1, ..., n$ identisch sind. Man mag sich die Realisierung der True-Scores und Observed-Scores einer Person also als unabhängige und identische Realisierungen aus einer "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1}, y_{\bullet 1}, ..., \tau_{\bullet m}, y_{\bullet m}) = \mathbb{P}(\tau_{\bullet 1}, ..., \tau_{\bullet m}) \prod_{j=1}^{m} \mathbb{P}(y_{\bullet j}|\tau_{\bullet j}) \tag{1.6}$$

vorstellen, wobei das Subskript $\bullet$ die Unspezifität dieser Verteilung bezüglich einer Person symbolisieren soll.

Betrachtet man den Spezialfall einer einzelnen Testmessung bei n Personen, so ergibt sich eine vereinfachte Form von Definition 1.2, auf die man häufig in der psychologischen Einführungsliteratur trifft. Nach Definition 1.2 gilt für $m = 1$

$$\mathbb{P}\left(\tau_{11}, y_{11}, ..., \tau_{n1}, y_{n1}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_{i1}) \mathbb{P}(y_{i1}|\tau_{i1}) \tag{1.7}$$

wobei nach Annahme von Definition 1.2 die gemeinsamen Marginalverteilungen $\mathbb{P}(\tau_{i1}, y_{i1})$ über Personden $i = 1, ..., n$ identisch sind. Wie oben kann man sich die observierten True- und Observed-Scores für eine Testmessung also als unabhängige Realisierungen der "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1})\mathbb{P}(y_{\bullet 1}|\tau_{\bullet 1}) \tag{1.8}$$

vorstellen. Verzichtet man nun noch auf die Subskripte $\bullet 1$, gelangt man zu folgender vereinfachter Definition.

Definition 1.3 (Vereinfachtes 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,

$$\mathbb{P}(\tau, y) = \mathbb{P}(\tau)\mathbb{P}(y|\tau) \tag{1.9}$$

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

•

Definition 1.3 hat gegenüber Definition 1.2 den Vorteil, dass weniger Zufallsvariablen und Indizes auftreten und auf die Redundanz der unterschiedlichen Bezeichnung vieler gleicher Verteilungen verzichtet werden kann. Im Sinne Frequentistischer Produktmodelle mag man bezüglich der Daten von n Personen hier auch

$$(\tau_1, y_1), ..., (\tau_n, y_n) \sim \mathbb{P}(\tau, y) \tag{1.10}$$

schreiben, wobei nur die $y_1, ..., y_n$ beobachtete, die $\tau_1, ..., \tau_n$ dagegen natürlich latente Zufallsvariablen sind. Weiterhin gilt, dass man viele wichtige Eigenschaften des Modells der Klassischen Testtheorie schon basierend auf den Eigenschaften von $\mathbb{P}(\tau, y)$ begründen kann, wie wir unten sehen werden. Generell ist das vereinfachte Modell der Klassischen Testtheorie einfacher zu handhaben als das Modell multipler Testmessungen. Problematisch wird es allerdings insbesondere, sobald mehrere Testmessungen (z.B. bei Abhängigkeitsbetrachtungen zwischen zwei Tests oder zwei Items eines Tests) ins Spiel kommen, wie es bei den interessanteren Aussagen der Klassischen Testtheorie grundsätzlich der Fall ist. Außerdem gilt natürlich auch, dass Schätzer der Modellparameter einerseits immer auf allen Observed-Score Zufallsvariablen $y_{1j}, ...y_{nj}$ beruhen und andererseits diese in Definition 1.3 überhaupt nicht auftreten. In anderen Worten ist Definition 1.3 für theoretische Betrachtungen oft einfacher zu behandeln als Definition 1.2, der Anwendung der Klassischen Testtheorie in der Fragebogendatenanalyse liegt in der Regel aber Definition 1.2 zugrunde. Wir werden in der Folge je nach Bedarf zwischen beiden Modellformulierungen hin und her wechseln, halten aber fest, dass die Modellformulierung in Sinne von Definition 1.2 unser Standardfall ist.

Eigenschaften des Modells multipler Testmessungen

Das Modell multipler Testmessungen nach Definition 1.2 hat zunächst eine Reihe von Eigenschaften bezüglich einer (und damit jeder) Testmessung, die für die Anwendung der Klassischen Testtheorie grundlegend sind.

1.1.0.1. Eigenschaften bezüglich einer Testmessungen

Wir fassen fünf dieser Eigenschaften bezüglich einer einzelnen Testmessung in folgendem Theorem zusammen.

Theorem 1.1 (Eigenschaften bezüglich einer Testmessung). *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) $\mathbb{V}(y_{ij}) = \mathbb{V}(\tau_{ij}) + \mathbb{V}(\varepsilon_{ij})$
(5) $\mathbb{C}(y_{ij}, \tau_{ij}) = \mathbb{V}(\tau_{ij})$

∘

Beweis. Zur Vereinfachung der Notation im Sinne des einfachen Modells der klassischen Testtheorie nach Definition 1.3 setzen wir zunächst für alle $i = 1, \dots, n$ und $j = 1, \dots n$

$$y := y_{ij} \text{ und } \tau := \tau_{ij} \text{ mit Ergebnisräumen } Y := Y_{ij} \text{ und } T := T_{ij}. \tag{1.11}$$

Weiterhin bezeichnen wir Werte von y mit y und Werte von τ mit t. Schließlich betrachten wir nur den diskreten Fall, setzen also die Existenz einer Wahrscheinlichkeitsmassefunktion $p : Y \times T \to [0, 1]$ der Form

$$p(t, \mathrm{y}) = p(\mathrm{y}|t)p(t) \tag{1.12}$$

voraus. Der kontinuierliche Fall oder gemischte Fall folgt dann jeweils analog.

(1) Es gilt

$$\begin{aligned}
\mathbb{E}(\varepsilon|\tau = t) &:= \mathbb{E}(y - \tau|\tau = t) \\
&= \sum_{\mathrm{y} \in Y} (\mathrm{y} - t)\, p(\mathrm{y}|t) \\
&= \sum_{\mathrm{y} \in Y} \mathrm{y} p(\mathrm{y}|t) - \sum_{\mathrm{y} \in Y} t p(\mathrm{y}|t) \\
&= \mathbb{E}(y|\tau = t) - t \sum_{\mathrm{y} \in Y} p(\mathrm{y}|t) \\
&= t - t \cdot 1 \\
&= 0.
\end{aligned} \tag{1.13}$$

(2) Es gilt

$$\begin{aligned}
\mathbb{E}(\varepsilon) &:= \mathbb{E}(y - \tau) \\
&= \sum_{t \in T} \sum_{\mathrm{y} \in Y} (\mathrm{y} - t)\, p(t, \mathrm{y}) \\
&= \sum_{t \in T} \sum_{\mathrm{y} \in Y} (\mathrm{y} - t)\, p(\mathrm{y}|t)p(t) \\
&= \sum_{t \in T} \sum_{\mathrm{y} \in Y} \mathrm{y} p(\mathrm{y}|t)p(t) - t p(\mathrm{y}|t)p(t) \\
&= \sum_{t \in T} \sum_{\mathrm{y} \in Y} p(t)\, (\mathrm{y} p(\mathrm{y}|t) - t p(\mathrm{y}|t)) \\
&= \sum_{t \in T} p(t) \left(\sum_{\mathrm{y} \in Y} \mathrm{y} p(\mathrm{y}|t) - t \sum_{\mathrm{y} \in Y} p(\mathrm{y}|t) \right) \\
&= \sum_{t \in T} p(t)\, (t - t \cdot 1) \\
&= \sum_{t \in T} p(t) \cdot 0 \\
&= 0.
\end{aligned} \tag{1.14}$$

(3) Es gilt

$$
\begin{aligned}
\mathbb{C}(\tau,\varepsilon) &= \mathbb{E}\left((\tau-\mathbb{E}(\tau))(\varepsilon-\mathbb{E}(\varepsilon))\right)\\
&= \mathbb{E}\left((\tau-\mathbb{E}(\tau))\varepsilon\right)\\
&= \sum_{t\in T}\sum_{\mathrm{y}\in Y}\left((t-\mathbb{E}(\tau))(\mathrm{y}-t)\right)p(t,\mathrm{y})\\
&= \sum_{t\in T}\sum_{\mathrm{y}\in Y}(t-\mathbb{E}(\tau))(\mathrm{y}-t)p(\mathrm{y}|t)p(t)\\
&= \sum_{t\in T}p(t)(t-\mathbb{E}(\tau))\sum_{\mathrm{y}\in Y}(\mathrm{y}-t)\,p(\mathrm{y}|t)\\
&= \sum_{t\in T}p(t)(t-\mathbb{E}(\tau))\sum_{\mathrm{y}\in Y}(\mathrm{y}p(\mathrm{y}|t)-tp(\mathrm{y}|t))\\
&= \sum_{t\in T}p(t)(t-\mathbb{E}(\tau))\left(\sum_{\mathrm{y}\in Y}\mathrm{y}p(\mathrm{y}|t)-t\sum_{\mathrm{y}\in Y}p(\mathrm{y}|t)\right)\\
&= \sum_{t\in T}p(t)(t-\mathbb{E}(\tau))\,(t-t\cdot 1)\\
&= \sum_{t\in T}p(t)(t-\mathbb{E}(\tau))\cdot 0\\
&= 0.
\end{aligned}
\tag{1.15}
$$

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

$$
\mathbb{V}(y)=\mathbb{V}(\tau+\varepsilon)=\mathbb{V}(\tau)+\mathbb{V}(\varepsilon)+2\mathbb{C}(\tau,\varepsilon)=\mathbb{V}(\tau)+\mathbb{V}(\varepsilon)+2\cdot 0=\mathbb{V}(\tau)+\mathbb{V}(\varepsilon) \tag{1.16}
$$

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

$$
\mathbb{E}(\varepsilon\tau)=\mathbb{E}(\tau\varepsilon)=\mathbb{C}(\tau,\varepsilon)+\mathbb{E}(\tau)\mathbb{E}(\varepsilon)=0+\mathbb{E}(\tau)\cdot 0=0 \tag{1.17}
$$

gilt

$$
\begin{aligned}
\mathbb{C}(y,\tau) &= \mathbb{E}(y\tau)-\mathbb{E}(y)\mathbb{E}(\tau)\\
&= \mathbb{E}((\tau+\varepsilon)\tau)-\mathbb{E}(\tau+\varepsilon)\mathbb{E}(\tau)\\
&= \mathbb{E}(\tau^2+\varepsilon\tau)-(\mathbb{E}(\tau)+\mathbb{E}(\varepsilon))\,\mathbb{E}(\tau)\\
&= \mathbb{E}(\tau^2)+\mathbb{E}(\varepsilon\tau)-\mathbb{E}(\tau)^2-\mathbb{E}(\varepsilon)\mathbb{E}(\tau)\\
&= \mathbb{E}(\tau^2)+\mathbb{E}(\varepsilon\tau)-\mathbb{E}(\tau)^2-0\cdot\mathbb{E}(\tau)\\
&= \mathbb{E}(\tau^2)+0-\mathbb{E}(\tau)^2-0\cdot\mathbb{E}(\tau)\\
&= \mathbb{E}(\tau^2)-\mathbb{E}(\tau)^2\\
&= \mathbb{V}(\tau).
\end{aligned}
\tag{1.18}
$$

□

Wie die Subskripte in Theorem 1.1 verdeutlichen, beziehen sich die Aussagen von Theorem 1.1 auf die Zufallsvariablen zur Modellierung der Daten einer Person i und einer Testmessung j und gelten gleichermaßen für alle $i=1,...,n$ und $j=1,...,m$. Aussage (1) von Theorem 1.1 betrifft den Erwartungswert des Error-Scores bedingt auf einem festen Wert des True-Scores, in diesem Fall ist der True-Score also keine Zufallsvariable und die Verteilung von Interesse $\mathbb{P}(\varepsilon_{ij}|\tau_{ij}=t_{ij})$. Aussagen (2) - (5) beziehen sich auf Eigenschaften der (gemeinsamen) Marginalverteilungen von y_{ij}, τ_{ij} und ε_{ij}. Im Sinne des einfachen Modells der Klassischen Testtheorie werden obige Eigenschaften oft auch als

(1) $\mathbb{E}(\varepsilon|\tau=t)=0$
(2) $\mathbb{E}(\varepsilon)=0$
(3) $\mathbb{C}(\tau,\varepsilon)=0$
(4) $\mathbb{V}(y)=\mathbb{V}(\tau)+\mathbb{V}(\varepsilon)$

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

geschrieben. Wir wollen diese Eigenschaften nun noch kurz kommentieren.

Spezieller Weise *folgt* für das Modell multipler Messungen, dass der bedingte Erwartungswert des Error-Scores $\mathbb{E}(\varepsilon|\tau = t)$ gleich Null ist aus der Definition des True-Scores. Dies steht im direkten Gegenentwurf zu typischen Annahmen über Messfehler,
die üblicherweiseeinen Messfehler mit einem (bedingten) Erwartungswert von Null *definieren*, da sie von Null verschiedene Beiträge zu einem Datenpunkt als Teil der durch sie repräsentierten Theorie konzeptualisieren (vgl. **?@sec-grenzwerte**). Weil im Modell multipler Testmessungen bedingte Error-Score Erwartungswert für jeden True-Score t gleich Null, folgt dann, dass auch der marginale Erwarungswert des Errorscores $\mathbb{E}(\varepsilon)$ gleich Null ist.

Die Tatsache, dass im Modell multipler Testmessungen gilt, dass die Kovarianz der True- und Error-Scores gleich Null ist besagt bekanntlich, dass hohe oder niedrige True-Scores nicht systematisch mit hohen oder niedrigen Error-Scores assoziiert. Die daraus folgende Tatsachen, dass die Observed-Score Varianz additiv in Beiträge der True-Score und der Error-Score Varianz zerlegt werden kann und dass die Kovarianz von Observed-Score und True-Score einer Testmessung der Varianz der True Scores enstpricht werden an späterer Stelle essentiell für Eigenschaften der Reliabilität von Testmessungen sein.

Im Folgenden wollen wir Theorem 1.1 noch an einem konkreten ersten Beispiel nach Lord & Novick (1968), Exercise 2.17 für ein Modell multipler Testmessungen veranschaulichen, wobei wir hierbei natürlich auf eine einzige Testmessung fokussieren.

Theorem 1.2 (Normalverteilungsbeispiel)**.** *Es sei $i = 1, ..., n$, $m := 1$ mit*

$$\mathbb{P}(\tau_{i1}) := N(\mu, 1) \text{ und } \mathbb{P}(y_{i1}|\tau_{i1}) := N(\tau_{i1}, 1) \tag{1.19}$$

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) $\mathbb{C}(\tau_{i1}, \varepsilon_{i1}) = 0$

∘

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

(1) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(y|\tau) = N(\tau, 1) \tag{1.20}$$

induzierte gemeinsame Verteilung von τ und y, wobei offenbar

$$\mathbb{P}(y|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 1 \text{ und } b := 0 \tag{1.21}$$

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

$$\begin{pmatrix} \tau \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ 1 \cdot \mu + 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \cdot 1 \\ 1 \cdot 1 & 1 + 1 \cdot 1 \cdot 1 \end{pmatrix}\right) = N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right) \tag{1.22}$$

Aus dem Theorem zu marginalen Normalverteilungen (LINK) 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

$$\varepsilon := y - t \text{ mit } y \sim N(t, 1) \tag{1.23}$$

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

$$\varepsilon \sim N\left(t - t, 1^2 \cdot 1\right) = N(0,1) \tag{1.24}$$

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

(3) und (4) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(\varepsilon|\tau) = N(0,1) \tag{1.25}$$

induzierte gemeinsame Verteilung von τ und ε, wobei offenbar

$$\mathbb{P}(\varepsilon|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 0 \text{ und } b := 0 \tag{1.26}$$

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

$$\begin{pmatrix}\tau \\ \varepsilon\end{pmatrix} \sim N\left(\begin{pmatrix}\mu \\ 0\cdot\mu + 0\end{pmatrix}, \begin{pmatrix}1 & 1\cdot 0 \\ 0\cdot 1 & 1 + 0\cdot 1\cdot 0\end{pmatrix}\right) = N\left(\begin{pmatrix}\mu \\ 0\end{pmatrix}, \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}\right) \tag{1.27}$$

Aus dem Theorem zu marginalen Normalverteilungen ((LINK)) ergeben sich dann durch Ablesen

$$\varepsilon \sim N(0,1) \text{ und } \mathbb{C}(\tau, \varepsilon) = 0. \tag{1.28}$$

□

Das Beispiel zeigt insbesondere, wie die Spezifika des Modells multipler Testmessungen der Klassischen Testtheorie äquivalent durch Definition eines üblichen probabilistischen Modells erzeugt werden können. Hier induzieren die Definition der marginalen Verteilung der latenten True-Score Variable und die Definition der bedingten Verteilung der observierten Observed-Score Variable zunächst eine eine gemeinsame Normalverteilung von τ_{i1} und y_{i1}. Die Definition der Error-Score Variable als Differenz zwischen Observed-Score und bedingtem True-Score Erwartungswert ergibt dann die bedingte Error-Score Verteilung. Diese und die Definitionen des Beispiels induzieren dann eine gemeinsame Normalverteilung von τ_{i1} und ε_{i1}. Die weiteren Eigenschaften im Sinne von Theorem 1.1 ergeben sich im Beispiel dann mit den Eigenschaften gemeinsamer Normalverteilungen.

SIMULATION UND VISUALISIERUNG

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(Y[,1],E[,1])                            # Kovarianzschätzung Observed-Score, True-Score
```

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

Eigenschaften bezüglich zweier Testmessungen

Bisher haben wir fünf Eigenschaften des Modells multipler Testmessungen kennengelernt, die für die True-, Observed-, und Error-Scores τ, y und ε einer (und damit jeder) Person i und einer Testmessung j gelten. Im Folgenden beschäftigen wir uns mit Eigenschaften des Modells multipler Testmessungen, die für die True-, Observed-, und Error-Scores $\tau_j, y_j, \varepsilon_j$ und $\tau_k, y_k, \varepsilon_k$ einer (und damit jeder) Person i hinsichtlich zweier Testmessungen j und k gelten. Wir fassen diese Eigenschaften, die manchmal als *lokale Unkorreliertheit* des Modells multipler Testmessungen bezeichnet werden, in folgende Theorem zusammen.

Theorem 1.3 (Eigenschaften bezüglich zweier Testmessungen)**.** *Gegeben sei das Modell multipler Testmessungen. Dann gelten für alle Personen $i = 1, ..., n$ und alle Testmessungen j und k mit $1 \leq j, k \leq m$ und $j \neq k$, dass*

(1) $\mathbb{C}(y_{ij}, y_{ik} | \tau_{ij} = t_{ij}, \tau_{ik} = t_{ik}) = 0$,
(2) $\mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik} | \tau_{ij} = t_{ij}, \tau_{ik} = t_{ik}) = 0$,
(3) $\mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik}) = 0$,
(4) $\mathbb{C}(\tau_{ij}, \varepsilon_{ik}) = 0$,
(5) $\mathbb{C}(y_{ij}, y_{ik}) = \mathbb{C}(\tau_{ij}, \tau_{ik})$.

∘

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

$$p(t_j, \mathrm{y}_j, t_k, y_k) = p(t_j, t_k) p(\mathrm{y}_j | t_j) p(\mathrm{y}_k | t_k) \tag{1.29}$$

und folglich auch der bedingten Wahrscheinlichkeitsmassefunktion

$$p(\mathrm{y}_j, y_k | t_j, t_k) = \frac{p(t_j, \mathrm{y}_j, t_k, y_k)}{p(t_j, t_k)} = \frac{p(t_j, t_k) p(\mathrm{y}_j | t_j) p(\mathrm{y}_k | t_k)}{p(t_j, t_k)} = p(\mathrm{y}_j | t_j) p(\mathrm{y}_k | t_k) \tag{1.30}$$

voraus. Der kontinuierliche Fall folgt dann wieder analog.

(1) Es gilt

$$\begin{aligned}
\mathbb{C}(y_j, y_k | \tau_j = t_j, \tau_k = t_k) &= \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - \mathbb{E}(y_j | \tau_j = t_j))(\mathrm{y}_k - \mathbb{E}(y_k | \tau_k = t_k)) p(\mathrm{y}_j | t_j) p(\mathrm{y}_k | t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j | t_j) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t_k) p(\mathrm{y}_k | t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j | t_j) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k | t_k) - t_k \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k | t_k) \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j | t_j) \left(t_k - t_k \cdot 1 \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j | t_j) \cdot 0 \\
&= 0.
\end{aligned} \tag{1.31}$$

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

$$\begin{aligned}
\mathbb{E}(\varepsilon_j\varepsilon_k|\tau_j = t_j, \tau_k = t_k) &= \mathbb{E}((y_j - \tau_j)(y_k - \tau_k)|\tau_j = t_j, \tau_k = t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - t_j)(\mathrm{y}_k - t_k) p(\mathrm{y}_j|t_j) p(\mathrm{y}_k|t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j|t_j) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t_k) p(\mathrm{y}_k|t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j|t_j) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t_k) - t_k \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k|t_k) \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j|t_j) \left(t_k - t_k \cdot 1 \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j) p(\mathrm{y}_j|t_j) \cdot 0 \\
&= 0
\end{aligned} \tag{1.32}$$

Mit dem Verschiebungssatz der bedingten Kovarianz LINK und Aussage (1) des Theorems zu den Ersten Eigenschaften des Modells multipler Messungen LINK folgt dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k|\tau_j = t_j, \tau_k = t_k) = \mathbb{E}(\varepsilon_j\varepsilon_k|\tau_j = t_j, \tau_k = t_k) - \mathbb{E}(\varepsilon_j|\tau_j = t_j)\mathbb{E}(\varepsilon_k|\tau_k = t_k) = 0 - 0 \cdot 0 = 0. \tag{1.33}$$

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

$$\begin{aligned}
\mathbb{E}(\varepsilon_j\varepsilon_k) &= \mathbb{E}((y_j - \tau_j)(y_k - \tau_k)) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - t_j)(\mathrm{y}_k - t_k) p(\mathrm{y}_j|t_j) p(\mathrm{y}_k|t_k) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - \mathbb{E}(y_j|\tau_j = t_j))(\mathrm{y}_k - \mathbb{E}(y_k|\tau_k = t_k)) p(\mathrm{y}_j|t_j) p(\mathrm{y}_k|t_k) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \mathbb{E}\left((y_j - \tau_j)(y_k - \tau_k)|\tau_j = t_j, \tau_j = t_j\right) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \mathbb{E}\left(\varepsilon_j\varepsilon_k|\tau_j = t_j, \tau_j = t_j\right) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} 0 \cdot p(t_j, t_k) \\
&= 0.
\end{aligned} \tag{1.34}$$

Mit dem Kovarianzverschiebungssatz LINK und Aussage (2) des Theorems zu den Ersten Eigenschaften des Modells multipler Testmessungen LINK ergibt sich dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = \mathbb{E}(\varepsilon_j\varepsilon_k) - \mathbb{E}(\varepsilon_j)\mathbb{E}(\varepsilon_k) = 0 - 0 \cdot 0 = 0. \tag{1.35}$$

(4) Mit dem Kovarianzverschiebungssatz LINK und Aussage (2) des Theorems zu den Ersten Eigenschaften

des Modells multipler Testmessungen LINK gilt

$$
\begin{aligned}
\mathbb{C}(\tau_j, \varepsilon_k) &= \mathbb{E}(\tau_j \varepsilon_k) - \mathbb{E}(\tau_j)\mathbb{E}(\varepsilon_k) \\
&= \mathbb{E}(\tau_j \varepsilon_k) - \mathbb{E}(\tau_j) \cdot 0 \\
&= \mathbb{E}(\tau_j (y_k - \tau_k)) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \sum_{\mathrm{y}_k \in Y_k} t_j (y_k - t_k) p(\mathrm{y}_k | t_k) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \sum_{\mathrm{y}_k \in Y_k} (y_k - t_k) p(\mathrm{y}_k | t_k) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \left(\sum_{\mathrm{y}_k \in Y_k} y_k p(\mathrm{y}_k | t_k) - t_k \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k | t_k) \right) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \left(t_k - t_k \cdot 1 \right) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \cdot 0 \\
&= 0.
\end{aligned}
\tag{1.36}
$$

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

$$
\mathbb{C}(\tau_k, \varepsilon_j) = \mathbb{C}(\varepsilon_j, \tau_k) = 0 \tag{1.37}
$$

gilt. Mit dem Theorem zu den Eigenschaften der Kovarianz aus Einheit (2) Theoretische Grundlagen und Aussage (3) gilt dann

$$
\begin{aligned}
\mathbb{C}(y_j, y_k) &= \mathbb{C}(\tau_j + \varepsilon_j, \tau_k + \varepsilon_k) \\
&= \mathbb{C}(\tau_j, \tau_k) + \mathbb{C}(\tau_j, \varepsilon_k) + \mathbb{C}(\varepsilon_j, \tau_k) + \mathbb{C}(\varepsilon_j, \varepsilon_k) \\
&= \mathbb{C}(\tau_j, \tau_k) + 0 + 0 + 0 \\
&= \mathbb{C}(\tau_j, \tau_k).
\end{aligned}
\tag{1.38}
$$

□

Im Sinne des vereinfachten Modells der Klassischen Testtheorie nach Definition 1.3 werden die Aussagen von Theorem 1.3 oft auch als

(1) $\mathbb{C}(y_j, y_k | \tau_j = t_j, \tau_k = t_k) = 0$
(2) $\mathbb{C}(\varepsilon_j, \varepsilon_k | \tau_j = t_j, \tau_k = t_k) = 0$
(3) $\mathbb{C}(\varepsilon_j, \varepsilon_k) = 0$
(4) $\mathbb{C}(\tau_j, \varepsilon_k) = 0$
(5) $\mathbb{C}(y_j, y_k) = \mathbb{C}(\tau_j, \tau_k)$

geschrieben. Die ersten beiden Aussagen von Theorem 1.3 besagen also, dass, bedingt auf den jeweiligen True Scores, die Kovarianzen von Observed-Scores sowie die Kovarianzen der Error-Scores einer Person zwischen zwei Testmessungen gleich Null sind. Für die Error-Scores gilt dies nach Aussage (3) von Theorem 1.3 auch im Sinne der unbedingten Marginalverteilung.Dies entspricht also der paarweisen Unabhängigkeit von Error-Scores über Testmessungen im Modell multipler Testmessungen. Weiterhin ist nach Aussage (4) die Kovarianz der True-Score Variable bei einer Testmessung mit der Error-Score Variable bei einer anderen Testmessung gleich Null. Schließlich gilt nach Aussage (5), dass die Kovarianz der observierbaren Observed-Scores zwischen zwei Testmessungen gleich der Kovarianz der latenten True-Scores ist.

Beispiel

Als erstes Beispiel für ein Modell multipler Testmessungen mit $m > 1$ setzen wir obiges Beispiel fort und betrachten den Fall zweier Testmessungen $j = 1, 2$. Für $i = 1, ..., n$ seien entsprechend

$$\mathbb{P}(\tau_{i1}, \tau_{i2}) = \mathbb{P}(\tau_{i2}|\tau_{i1})\mathbb{P}(\tau_{i1}) \tag{1.39}$$

mit

$$\mathbb{P}(\tau_{i1}) := N(1,1) \text{ und } \mathbb{P}(\tau_{i2}|\tau_{i1}) := N(\tau_{i1}+1, 1) \tag{1.40}$$

Die Verteilung des True-Score von Person i in Testmessung $j = 2$ hängt in diesem Beispiel also explizit von der Verteilung des True-Scores von Person i in Testmessung $j = 1$ ab. Weiterhin seien

$$\mathbb{P}(y_{i1}|\tau_{i1}) := N(\tau_{i1}, 1) \text{ und } \mathbb{P}(y_{i2}|\tau_{i2}) := N(\tau_{i2}, 2) \tag{1.41}$$

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

SIMULATION UND VISUALISIERUNG

```
n           = 1e5                               # Personenanzahl
m           = 2                                 # 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
    T[i,1]  = rnorm(1,1,1)                      # True-Score Realisierung     für j = 1
    Y[i,1]  = rnorm(1,T[i,1],1)                 # Observed-Score Realisierung für j = 1
    E[i,1]  = Y[i,1] - T[i,1]                   # Error-Score Realisierung    für j = 1
    T[i,2]  = rnorm(1,T[i,1] + 1,.5)            # True-Score Realisierung     für j = 2
    Y[i,2]  = rnorm(1,T[i,2],.5)                # Observed-Score Realisierung für j = 2
    E[i,2]  = Y[i,2] - T[i,2]}                  # Error-Score Realisierung    für j = 2
c_hat_e1_e2 = cov(E[,1],E[,2])                  # Kovarianzschätzung Error-Score 1, Error-Score 2
c_hat_t1_e2 = cov(T[,1],E[,2])                  # Kovarianzschätzung True-Score 1, Error-Score 2
c_hat_o1_o2 = cov(Y[,1],Y[,2])                  # Kovarianzschätzung Observed-Score 1, Observed-Score 2
c_hat_t1_t2 = cov(T[,1],T[,2])                  # Kovarianzschätzung True-Score 1, True-Score 2
```

$n = 500$ Realisierungen

1.2. 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 $i = 1, ..., n$

$$\mathbb{P}(\tau_{i1}, ..., \tau_{im}) \neq \prod_{j=1}^{m} \mathbb{P}(\tau_{ij}). \tag{1.42}$$

Andererseits haben wir die Form möglicher Abhängigkeiten zwischen den True-Scores $\tau_{i1}, ..., \tau_{im}$ bislang nicht genauer spezifiziert. Die Klassische Testtheorie betrachtet in dieser Hinsicht einige Spezialfälle, die sich allgemein durch funktionale Abhängigkeiten zwischen τ_{i1} und $\tau_{i2}, ..., \tau_{im}$ der Form

$$\tau_{ij} = f(\tau_{i1}) \text{ für } f : \mathbb{R} \to \mathbb{R} \text{ und } j = 2, ..., m. \tag{1.43}$$

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

$$t_{ij} = \mathrm{id}_{\mathbb{R}}\left(t_{i1}\right) = t_{i1} \text{ für } j = 2, ..., m, \tag{1.44}$$

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

$$\tau_{ij} = f(\tau_{i1}) = a\tau_{i1} + b \text{ für } a, b \in \mathbb{R} \text{ und } j = 2, ..., m. \tag{1.45}$$

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

Definition 1.4 (Modell paralleler Testmessungen)**.** Für $i = 1, ..., n$ und $j = 1, ..., m$ seien τ_i eine Zufallsvariable, die den *True-Score* der $i$ten Person in jeder Testmessung $j = 1, ..., m$ modelliere, y_{ij} eine Zufallsvariable, die den *Observed-Score* der $i$ten Person in der $j$ten Testmessung modelliere und $\varepsilon_{ij} := y_{ij} - \tau_i$ die Zufallsvariable, die den *Error-Score* der $i$ten Person in der $j$ten Testmessung modelliere. Dann heißt die gemeinsame Verteilung der τ_i und y_{ij} mit den Faktorisierungseigenschaften

$$\mathbb{P}\left(\tau_1, y_{11}, ...y_{1m}, ..., \tau_n, y_{n1}, ..., y_{nm}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_i) \prod_{j=1}^{m} \mathbb{P}(y_{ij}|\tau_i) \tag{1.46}$$

das *Modell paralleler Testmessungen*, wenn gilt, dass

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

•

Insbesondere werden also im Modell paralleler Testmessungen die Werte des True-Scores einer Person werden über Testmessungen hinweg als identisch angenommen. Damit geht dann die Observed-Score Varianz einer Person zwischen zwei Testmessungen allein auf die Propensitätsverteilung zurück. Aus generativer Sichtweise entstehen Werte der Observed Scores also wie folgt: Zunächst wird für die $i$te Person i und Testmessungen $j = 1, ..., m$ ein True-Score t_i von gemäß $\mathbb{P}(\tau_i)$ realisiert. Dann wird für die $i$te und die Testmessung $j = 1, ..., m$ ein Observed-Score y_{ij} anhand von $\mathbb{P}(y_{ij}|\tau_i = t_i)$ realisiert. Betrachtet man nur eine einzige Testmessung, so hat das Modell paralleler Testmessungen natürlich die gleiche Form wie das Modell multipler Testmessungen. Damit gilt dann aber auch Theorem 1.1 analog für das Modell paralleler Testmessungen. Betrachtet man allerdings mehr als eine Testmessung, so ergeben sich für das Modell paralleler Testmessungen speziellere Eigenschaften, die wir in folgendem Theorem festhalten

1.2.1. Eigenschaften des Modells paralleler Testmessungen

Theorem 1.4 (Eigenschaften des Modells paralleler Testmessungen)**.** *Gegeben sei das Modell paralleler Testmessungen. Dann gelten für alle $i = 1, ..., n$ und alle j, k mit $1 \le j, k \le m$ und $j \ne k$, 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) $\mathbb{C}(y_{ij}, y_{ik}) = \mathbb{V}(\tau_i)$

∘

Beweis. Zum Beweis setzen zur Vereinfachung der Notation zunächst

$$y_j := y_{ij}, y_k := y_{ik}, \tau := \tau_i, \mathrm{y}_j := \mathrm{y}_{ij}, \mathrm{y}_k := \mathrm{y}_{ik}, t := t_i, Y_j := Y_{ij}, Y_k := Y_{ik} \text{ und } T := T_i \tag{1.47}$$

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

$$p(t, \mathrm{y}_j, \mathrm{y}_k) = p(\mathrm{y}_j|t)p(\mathrm{y}_k|t)p(t) \tag{1.48}$$

voraus. Der kontinuierliche Fall folgt dann analog.

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

$$\mathbb{E}(y_j) = \sum_{t\in T}\sum_{\mathrm{y}_j\in Y_j} \mathrm{y}_j p(\mathrm{y}_j|t)p(t) = \sum_{t\in T}\mathbb{E}(y_j|\tau = t)p(t) = \sum_{t\in T}\sum_{\mathrm{y}_k\in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t)p(t) = \mathbb{E}(y_k). \tag{1.49}$$

(2) Mit der Darstellung der Varianz LINK ergibt sich

$$\mathbb{V}(y_j) = \mathbb{V}\left(\mathbb{E}(y_j|\tau)\right) + \mathbb{E}\left(\mathbb{V}(y_j|\tau)\right) = \mathbb{V}\left(\mathbb{E}(y_k|\tau)\right) + \mathbb{E}\left(\mathbb{V}(y_k|\tau)\right) = \mathbb{V}(y_k). \tag{1.50}$$

(3) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j\varepsilon_k)$. Mit Aussage (2) von Theorem 1.1 gilt dann

$$\begin{aligned}
\mathbb{E}(\varepsilon_j\varepsilon_k) &:= \mathbb{E}((y_j-\tau)(y_k-\tau))\\
&= \sum_{t\in T}\sum_{\mathrm{y}_j\in Y_j}\sum_{\mathrm{y}_k\in Y_k}(\mathrm{y}_j - t)(\mathrm{y}_k - t)p(t)p(\mathrm{y}_j|t)p(\mathrm{y}_k|t)\\
&= \sum_{t\in T}p(t)\sum_{\mathrm{y}_j\in Y_j}\sum_{\mathrm{y}_k\in Y_k}(\mathrm{y}_j - t)(\mathrm{y}_k - t)p(\mathrm{y}_j|t)p(\mathrm{y}_k|t)\\
&= \sum_{t\in T}p(t)\sum_{\mathrm{y}_j\in Y_j}(\mathrm{y}_j - t)p(\mathrm{y}_j|t)\sum_{\mathrm{y}_k\in Y_k}(\mathrm{y}_k - t)p(\mathrm{y}_k|t)\\
&= \sum_{t\in T}p(t)\left(\sum_{\mathrm{y}_j\in Y_j}\mathrm{y}_j p(\mathrm{y}_j|t) - t\sum_{\mathrm{y}_j\in Y_j}p(\mathrm{y}_j|t)\right)\left(\sum_{\mathrm{y}_k\in Y_k}\mathrm{y}_k p(\mathrm{y}_k|t) - t\sum_{\mathrm{y}_k\in Y_k}p(\mathrm{y}_k|t)\right)\\
&= \sum_{t\in T}p(t)\,(t-t)\,(t-t)\\
&= \sum_{t\in T}p(t)\cdot 0\cdot 0\\
&= 0.
\end{aligned} \tag{1.51}$$

Mit dem Kovarianzverschiebungssatz LINK und wiederrum mit Aussage (2) von Theorem 1.1 folgt dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = \mathbb{E}(\varepsilon_j\varepsilon_k) - \mathbb{E}(\varepsilon_j)\mathbb{E}(\varepsilon_k) = 0 - 0\cdot 0 = 0 \tag{1.52}$$

(4) Mit dem Kovarianzverschiebungssatz LINK und Aussage (2) von Theorem 1.1

$$
\begin{aligned}
\mathbb{C}(\tau, \varepsilon_k) &= \mathbb{E}(\tau\varepsilon_k) - \mathbb{E}(\tau)\mathbb{E}(\varepsilon_k) \\
&= \mathbb{E}(\tau\varepsilon_k) - \mathbb{E}(\tau) \cdot 0 \\
&= \mathbb{E}(\tau(y_k - \tau)) \\
&= \sum_{t \in T} \sum_{\mathrm{y}_k \in Y_k} t(\mathrm{y}_k - t)p(t)p(\mathrm{y}_k|t) \\
&= \sum_{t \in T} tp(t) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t)p(\mathrm{y}_k|t) \\
&= \sum_{t \in T} tp(t) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t) - t \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k|t) \right) \\
&= \sum_{t \in T} tp(t)\,(t - t) \\
&= \sum_{t \in T} tp(t) \cdot 0 \\
&= 0.
\end{aligned}
\tag{1.53}
$$

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

$$
\mathbb{C}(\tau, \varepsilon_j) = \mathbb{C}(\varepsilon_j, \tau) = 0 \tag{1.54}
$$

gilt. Mit dem Theorem zu den Eigenschaften der Kovarianz LINK Aussage (1) des Theorems gilt dann

$$
\begin{aligned}
\mathbb{C}(y_j, y_k) &= \mathbb{C}(\tau + \varepsilon_j, \tau + \varepsilon_k) \\
&= \mathbb{C}(\tau, \tau) + \mathbb{C}(\tau, \varepsilon_k) + \mathbb{C}(\varepsilon_j, \tau) + \mathbb{C}(\varepsilon_j, \varepsilon_k) \\
&= \mathbb{C}(\tau, \tau) + 0 + 0 + 0 \\
&= \mathbb{C}(\tau, \tau) \\
&= \mathbb{V}(\tau).
\end{aligned}
\tag{1.55}
$$

□

Aussage (1) des Theorems besagt, dass bei Paralleltestmessungen alle Erwartungswerte der Observed Scores identisch sind und Aussage (2) besagt, dass bei Paralleltestmessungen alle Varianzen der Observed Scores identisch sind. Die Aussagen (3) und (4) sind analog zu den lokalen Unkorreliertheitseigenschaften des Modells multipler Testmessungen. Aussage (5) besagt insbesondere, dass $\mathbb{C}(y_{ij}, y_{ik})$ für beliebe j und k identisch zu $\mathbb{V}(\tau_i)$ sind. Mit 2 sind im Modell paralleler Testmessungen also alle paarweisen Korrelationen verschiedener Testmessungen identisch.

Beispiel

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

$$
\mathbb{P}(\tau_i) = N(1, 1) \text{ und } \mathbb{P}(y_{i1}|\tau_i) := \mathbb{P}(y_{i2}|\tau_i) := N(\tau_i, 1) \tag{1.56}
$$

Für Person i gibt es also nur eine True-Score Zufallsvariable für alle Testmessungen und die Propensitätsverteilungen unterscheiden sich zwischen Testmessungen nicht.

SIMULATION UND VISUALISIERUNG

```
n           = 1e5                              # Personenanzahl
m           = 2                                # Testmessungsanzahl
T           = matrix(rep(NaN, n)  , 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){                                 # Personeniterationen
    T[i]  = rnorm(1,1,1)                       # True-Score Realisierung für j = 1,2
    for(j in 1:m){                             # Testmessungsiterationen
        Y[i,j]  = rnorm(1,T[i],1)              # Observed-Score Realisierung f
        E[i,j]  = Y[i,j] - T[i]}}              # Error-Score Realisierung
e_hat_o1_o2 = apply(Y, 2, mean)                # Erwartungswertschätzung Observed-Score 1, Observed-Score 2
v_hat_o1_o2 = apply(Y, 2, var)                 # Varianzzschätzung Observed-Score 1, Observed-Score 2
c_hat_e1_e2 = cov(E[,1],E[,2])                 # Kovarianzzschätzung Error-Score 1, Error-Score 2
c_hat_t_e2  = cov(T    ,E[,2])                 # Kovarianzzschätzung True-Score 1, Error-Score 2
c_hat_o1_o2 = cov(Y[,1],Y[,2])                 # Kovarianzzschätzung Observed-Score 1, Observed-Score 2
v_hat_t     = var(T)                           # Varianzschätzung True-Score
```

1.3. Reliabilität

Die *Reliabilität* einesr Testmessung ist das zentrale Konzept der Klassischen Testtheorie. Wir folgen hier dem Ansatz nach HISTORISCHE REFERENZ, wonach die Reliabilität einer Testmessung als quadrierte Korrelation von Observed- und True-Score definiert ist. Basierend auf den Eigenschaften des Modells multipler Testmessungen ergeben sich dann verschiedene Möglichkeiten, diese Korrelation darzustellen und damit verschiedene Möglichkeiten, die Reliabilität einer Testmessung zu interpretieren. Allerdings ergibt sich dabei keine Möglichkeit, die Reliabilität von Testmessungen empirisch zu messen. Zentral ist dann die Übertragung des Konzepts der Reliabilität in das Modell paralleler Testmessungen. Die so definierte *Paralleltestreliabilität*, ist dann empirisch schätzbar. Wir beginnen zunächst mit der Definition der Reliabilität einer Testmessung im Modell multipler Testmessungen.

Definition 1.5 (Reliabilität einer Testmessung)**.** Gegeben sei das Modell multipler Testmessungen für eine beliebige Testmessung j mit $1 \le j \le m$,

$$\mathbb{P}(\tau_{1j}, y_{1j}..., \tau_{nj}, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(\tau_{ij}, y_{ij}) := \prod_{i=1}^{n} \mathbb{P}(y_{ij}|\tau_{ij})\mathbb{P}(\tau_{ij}) \tag{1.57}$$

wobei nach Definition des Modells gilt, dass

$$\mathbb{P}(\tau_{1j}, y_{1j}) = \cdots = \mathbb{P}(\tau_{nj}, y_{nj}). \tag{1.58}$$

Die *Reliabilität der Testmessung* j ist dann definiert als

$$\mathrm{R}_j := \rho(y_{ij}, \tau_{ij})^2 \text{ für ein beliebiges } 1 \le i \le n. \tag{1.59}$$

•

Vor dem Hintergrund des vereinfachen Modells der Klassischen Testtheorie Definition 1.3 schreibt man auch

$$\mathrm{R} = \rho(y, \tau)^2 \tag{1.60}$$

Per Definition ist die Reliabilität eine Testmessung also die quadrierte Korrelation von Observed- und True-Score. Mit

$$-1 \le \rho(y_{ij}, \tau_{ij}) \le 1 \tag{1.61}$$

folgt direkt, dass für die Reliailität einer Testmessung gilt, dass

$$0 \le \mathrm{R}_j \le 1 \tag{1.62}$$

Die Aussage $\mathrm{R}_j = 0$ impliziert dann $\rho(y_{ij}, \tau_{ij}) = 0$, also die linear-affine Unabhängigkeit von Observed-Score und True-Score (vgl. LINK) und damit den Umstand, dasss der Observed-Score hinsichtlich des True- Scores nicht aussagekräftig ist. $\mathrm{R}_j = 1$ dagegen impliziert $\rho(y_{ij}, \tau_{ij}) = \pm 1$, also die deterministische linear-affine Abhängigkeit von Observed-Score und True-Score und damit, dass der Observed-Score hinsichtlich des True-Scores vollständig aussagekräftig ist.

Folgendes Theorem zeigt weitere Möglichkeiten auf, die Reliabilität einer Testmessung äquivalent zu formulieren und damit zu interpretieren.

Theorem 1.5 (Eigenschaften der Reliabilität einer Testmessung)**.** *Gegeben sei das Modell multipler Testmessungen für eine beliebige Testmessung j mit $1 \le j \le m$,*

$$\mathbb{P}(\tau_{1j}, y_{nj} ..., \tau_{nj}, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(y_{ij}|\tau_{ij})\mathbb{P}(\tau_{ij}). \tag{1.63}$$

Dann gelten für die Reliabilät R_j der Testmessung j, dass

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

∘

Beweis. (1) Mit Aussage (5) von Theorem 1.1 gilt

$$\mathrm{R}_j = \rho(y_{ij}, \tau_{ij})^2 = \left(\frac{\mathbb{C}(y_{ij}, \tau_{ij})}{\mathbb{S}(y_{ij})\mathbb{S}(\tau_{ij})}\right)^2 = \frac{\mathbb{C}(y_{ij}, \tau_{ij})^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_{ij})} = \frac{\mathbb{V}(\tau_{ij})^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_{ij})} = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})}. \tag{1.64}$$

(2) Mit Aussage (4) von Theorem 1.1 gilt dann weiter

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij}) - \mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij})}{\mathbb{V}(y_{ij})} - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}. \tag{1.65}$$

□

Da τ_{ij} und ε_{ij} weiterhin latent und damit nur indirekt beobachtbar sind, sind die Darstellungen nach Theorem 1.5 nur von theoretischem Interesse. Die erste Aussage besagt dabei insbesondere, dass die Reliabilität einer Testmessung der Anteil der True-Score Varianz an der Observed-Score Varianz ist,

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(\tau_{ij}) + \mathbb{V}(\varepsilon_{ij})}. \tag{1.66}$$

Gilt dabei, dass $\mathbb{V}(\tau_{ij}) = 0$, so ist auch $\mathrm{R}_j = 0$, ist dagegen $\mathbb{V}(\varepsilon_{ij}) = 0$ so ist $\mathrm{R}_j = 1$. Eine Reliabilität $\mathrm{R}_j > 0$ impliziert also immer eine von Null verschiedene True-Score Varianz.

Paralleltestreliabilität

Um das Konzept der Reliabilität nun anhand empirischer Daten von Observed-Scores schätzbar zu machen, bedarf es seiner Übertragung in das Modell paralleler Testmessungen. Wir definieren daher die Reliabilität einer Paralleltestmessung explizit wie folgt.

Definition 1.6 (Reliabilität einer Paralleltestmessung)**.** Gegeben sei das Modell paralleler Testmessungen für eine beliebige Testmessung j mit $1 \le j \le m$,

$$\mathbb{P}(\tau_1, y_{1j}..., \tau_n, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(\tau_i, y_{ij}) = \prod_{i=1}^{n} \mathbb{P}(\tau_i)\mathbb{P}(y_{ij}|\tau_i) \tag{1.67}$$

wobei nach Definition des Modells gilt, dass

$$\mathbb{P}(\tau_1, y_{1j}) = \cdots = \mathbb{P}(\tau_n, y_{nj}). \tag{1.68}$$

Die *Reliabilität der Paralleltestmessung* j ist dann definiert als

$$\mathrm{R}_j := \rho(y_{ij}, \tau_i)^2 \text{ für ein beliebiges } 1 \le i \le n. \tag{1.69}$$

•

Vor dem Hintergrund des vereinfachten Modells der Klassischen Testtheorie nach Definition 1.3 schreibt man auch hier

$$\mathrm{R} = \rho(y, \tau)^2. \tag{1.70}$$

Theorem 1.6 (Paralleltestreliabilität)**.** *Gegeben sei das Modell der paralleler Testmessungen*

$$\mathbb{P}(\tau_1, y_{1j}..., \tau_n, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(\tau_i, y_{ij}) = \prod_{i=1}^{n} \mathbb{P}(\tau_i)\mathbb{P}(y_{ij}|\tau_i) \tag{1.71}$$

Dann gelten

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

∘

Beweis. (1) Mit Aussage (5) von Theorem 1.1 gilt

$$\mathrm{R}_j = \rho(y_{ij}, \tau_i)^2 = \left(\frac{\mathbb{C}(y_{ij}, \tau_i)}{\mathbb{S}(y_{ij})\mathbb{S}(\tau_i)}\right)^2 = \frac{\mathbb{C}(y_{ij}, \tau_i)^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_i)} = \frac{\mathbb{V}(\tau_i)^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_i)} = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})}. \tag{1.72}$$

(2) Mit Aussage (4) von Theorem 1.1 gilt dann weiter

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij}) - \mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij})}{\mathbb{V}(y_{ij})} - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}. \tag{1.73}$$

(3) Mit Aussagen (2) und (5) von Theorem 1.3 gilt dann weiter

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ij})}\sqrt{\mathbb{V}(y_{ij})}} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ij})}\sqrt{\mathbb{V}(y_{ik})}} = \rho(y_{ij}, y_{ik}) \tag{1.74}$$

und dass ebenso

$$\mathrm{R}_k = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ik})} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ik})}\sqrt{\mathbb{V}(y_{ik})}} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ij})}\sqrt{\mathbb{V}(y_{ik})}} = \rho(y_{ij}, y_{ik}). \tag{1.75}$$

□

Aussagen (1) und (2) sind von **?@thm-reliabilität-einer-paralleltestmessung** analog zum Theorem der Reliabilität multipler Testmessungen. Aussage (3) des Theorems begründet das empirische Vorgehen der Reliabilitätsschätzung mithilfe von Parallel- oder Retestverfahren. Zur Bestimmung der Reliabilität eines Tests nutzt man entspreched eine Schätzung der Korrelation zweier paralleler Testmessungen, wobei die Klassische Testtheorie dieses Vorgehen vor der Annahme von von latenten True- und Error-Scores begründet. Vereinfacht wird Aussage (3) oft dadurch ausgedrückt, dass man sagt, dass "die Korrelation paralleler Testmessungen gleich ihrer Reliabilität" ist.

Beispiel

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

$$p(t) = N(t; 0, \sigma_\tau^2) \text{ und } p(\mathrm{y}_1|t) := N(\mathrm{y}_1; t, \sigma_\varepsilon^2) \text{ und } p(\mathrm{y}_2|t) := N(\mathrm{y}_2; t, \sigma_\varepsilon^2) \tag{1.76}$$

Für Person i gibt es also nur eine True-Score Zufallsvariable für alle Testmessungen und die Propensitätsverteilungen unterscheiden sich zwischen Testmessungen nicht.

Dann gilt zunächst mit dem Theorem zu gemeinsamen Normalverteilungen mit $A := 1$ und $b := 0$

$$p(t)p(\mathrm{y}_1|t) = p(t, \mathrm{y}_1) = N\left(\begin{pmatrix} t \\ \mathrm{y}_1 \end{pmatrix}; \begin{pmatrix} \mu_\tau \\ \mu_\tau \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \tag{1.77}$$

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

$$p(t, \mathrm{y}_1)p(\mathrm{y}_2|t) = p(t, \mathrm{y}_1, \mathrm{y}_2) = N\left(\begin{pmatrix} t \\ \mathrm{y}_1 \\ \mathrm{y}_2 \end{pmatrix}; \begin{pmatrix} \mu_\tau \\ \mu_\tau \\ \mu_\tau \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \tag{1.78}$$

Damit gilt dann durch Ablesen am Kovarianzmatrixparameter von $p(t, \mathrm{y}_1, \mathrm{y}_2)$, dass

$$\rho(y_1, y_2) = \frac{\mathbb{C}(y_1, y_2)}{\sqrt{\mathbb{V}(y_1)}\sqrt{\mathbb{V}(y_2)}} = \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2}\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2}} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2} \tag{1.79}$$

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

$$\mathrm{R}_j = \rho(y_j, \tau)^2 = \left(\frac{\mathbb{C}(y_j, \tau)}{\mathbb{S}(y_j)\mathbb{S}(\tau)}\right)^2 = \frac{\mathbb{C}(y_j, \tau)^2}{\mathbb{V}(y_j)\mathbb{V}(\tau)} = \frac{\left(\sigma_\tau^2\right)^2}{(\sigma_\tau^2 + \sigma_\varepsilon^2)\sigma_\tau^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2} = \rho(y_1, y_2). \tag{1.80}$$

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

$$\mathrm{R}_j = \rho(y_j, \tau)^2 = \rho(y_1, y_2) = \frac{1.0}{1.0 + 0.2} \approx 0.83. \tag{1.81}$$

Paralleltestreliabilitätsschätung

Wie üblich wird die Korrelation zweier paralleler Testmessungen durch eine Stichprobenkorrelation geschätzt. Wir formulieren dies für das aktuelle Szenario in mithilfe folgender Definition.

Definition 1.7 (Paralleltestreliabilitätsschätzer)**.** Gegeben sei das Modell paralleler Testmessungen für n Personen und zwei Testmessungen $j = 1, 2$,

$$\mathbb{P}(\tau_1, y_{11}, y_{12}, ..., \tau_n, y_{n1}, y_{n2}) = \prod_{i=1}^{n} \mathbb{P}(\tau_i)\mathbb{P}(y_{i1}|\tau_i)\mathbb{P}(y_{i2}|\tau_i) \tag{1.82}$$

Dann wird der mit den Stichprobenmitteln

$$\bar{y}_1 := \frac{1}{n}\sum_{i=1}^{n} y_{i1} \text{ und } \bar{y}_2 := \frac{1}{n}\sum_{i=1}^{n} y_{i2} \tag{1.83}$$

definierte Stichprobenkorrelationskoeffizient

$$r_{12} := \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_{i2} - \bar{y}_2)^2}} \tag{1.84}$$

Paralleltestreliabilitätsschätzer genannt.

•

Bekanntlich bietet **R** mit der `cor()` Funktion eine Möglichkeit, Stichprobenkorrelationskoeffizienten zu berechnen. Wir demonstrieren dies an einem Simulationsbeispiel dazu seien

$$p(t_i) = N(t_i; 0, \sigma_\tau^2) \text{ und } p(y_{ij}|t_i) := N(y_{ij}; t_i, \sigma_\varepsilon^2) \tag{1.85}$$

für $j = 1, 2$ mit $\sigma_\tau^2 := 1.0$ und $\sigma_\varepsilon^2 := 0.2$ für $i = 1, ..., 30$.

```
n           = 30                                          # Personenanzahl
m           = 2                                           # Testmessungsanzahl
sigsqr_tau  = 1                                           # True-Score Varianz
sigsqr_eps  = .2                                          # Observed-Score-Varianz
R_12        = sigsqr_tau/(sigsqr_tau+sigsqr_eps)          # Paralleltestreliabilität
T           = matrix(rep(NaN, n)  , 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){                                            # Personeniterationen
    T[i]  = rnorm(1,1,sqrt(sigsqr_tau))                   # True-Score Realisierung für j = 1,2
    for(j in 1:m){                                        # Testmessungsiterationen
        Y[i,j]  = rnorm(1,T[i],sqrt(sigsqr_eps))          # Observed-Score Realisierung f
        E[i,j]  = Y[i,j] - T[i]}}                         # Error-Score Realisierung
r_12      = cor(Y[,1],Y[,2])                              # Paralleltestreliabilitätsschätzer
```

```
cat("Paralleltestreliabilität R_12           : ", round(R_12,4),
    "\nParalleltestrelibabilitätsschätzer  r_12 : ",  round(r_12,4))
```

```
Paralleltestreliabilität R_12           :  0.8333
Paralleltestrelibabilitätsschätzer  r_12 :  0.8892
```

Um ein Konfidenzintervall für die Paralleltestreliabilät anzugeben, ist eine Annahme zur empirischen Verteilung des Reliabilitätsschätzers erforderlich. Dazu nutzt man üblicherweise folgende asymptotische Verteilungsaussage.

Theorem 1.7 (Fishertransformation des Paralleltestreliabilitätsschätzers)**.** *Gegeben sei das Modell paralleler Testmessungen für n Personen mit Paralleltestreliabilität R_j. Für zwei Testmessungen $j = 1,2$ sei r_{12} der Paralleltestreliabilitätsschätzer. Dann gilt, dass*

$$\tilde{r}_{12} := \frac{1}{2}\ln\left(\frac{1+r_{12}}{1-r_{12}}\right) \tag{1.86}$$

asymptotisch normalverteilt ist nach

$$\tilde{r}_{12} \overset{a}{\sim} N\left(R_j, (n-3)^{-1}\right) \tag{1.87}$$

∘

Wir verzichten auf einen Beweis und verweisen für eine ausführliche Darstellung auf Johnson et al. (1994), Kapitel 32. Dabei wird die Transformation eines Stichprobenkorrelationskoeffizienten anhand von

$$\tilde{r} := \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{1.88}$$

wird nach Fisher (1925) als *Fisher-Transformation* bezeichnet.

SIMULATION UND VISUALISIERUNG

```
nsim                = 1e4                                    # Realisierungsanzahl
n                   = 30                                     # Personenanzahl
m                   = 2                                      # Testmessungsanzahl
sigsqr_tau          = 1                                      # True-Score Varianz
sigsqr_eps          = .2                                     # Observed-Score-Varianz
R_12                = sigsqr_tau/(sigsqr_tau+sigsqr_eps)     # Paralletestreliabilität
tilde_r_12          = rep(NaN, nsim)                         # Fisher Transformation von r_12 Array
for(s in 1:nsim){                                            # Simulationsiterationen
    T               = matrix(rep(NaN, n)  , 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){                                           # Personeniterationen
        T[i]        = rnorm(1,1,sqrt(sigsqr_tau))            # True-Score Realisierung für j = 1,2
        for(j in 1:m){                                       # Testmessungsiterationen
            Y[i,j]  = rnorm(1,T[i],sqrt(sigsqr_eps))         # Observed-Score Realisierung f
            E[i,j]  = Y[i,j] - T[i]}}                        # Error-Score Realisierung
    r_12            = cor(Y[,1],Y[,2])                       # Paralleltestreliabilitätsschätzer
    tilde_r_12[s]   = 1/2*log((1+r_12)/(1-r_12))             # Fisher Transformation von r_12
}
```

KONFIDENZINTERVALL FÜR RELIABILITÄTSSCHÄTZER.

1.4. Interne Konsistenz

m-Komponententestmodelle

In den vorherigen Abschnitten bezeichnete y_{ij} die Zufallsvariable zur Modellierung des Observed-Scores der $j$ten Testmessung der $i$ten Person, wobei wir m Testmessungen $j = 1, ..., m$ und n Personen $i = 1, ..., n$ angenommen haben. Wir haben dabei aber zunächst offen gelassen, ob mit der *jten Testmessung* das summative Ergebnis eines *Tests* oder ein einzelnes *Item* eines Tests gemeint ist um die Theorie für beide Anwendungsfälle zu entwickeln. Als Grundlage der Bestimmung der inneren Konsistenz eines Tests identifizieren wir in diesem Abschnitt nun die $j$te Testmessung nun mit dem $j$ten *Item eines Tests*. y_{ij} bezeichnet also im Folgenden die Zufallsvariable zur Modellierung des Observed-Scores des $j$ten Items der $i$ten Person in einem Test, wobei wir weiterhin m Items $j = 1, ..., m$ und n Personen $i = 1, ..., n$ annehmen. Weiterhin gehen wir in diesem Zusammenhang davon aus, dass für jede Person ein *Gesamt-Observed-Score* durch Summation über die Items eines Test gebildet wird. Die Zufallsvariable zur Modellierung dieses Gesamt-Observed-Scores bezeichnen wir entsprechend mit

$$y_i := \sum_{j=1}^{m} y_{ij}. \tag{1.89}$$

Wir fassen diese Vorüberlegungen mit folgender Definition des *m-Komponententestmodells* zusammen.

Definition 1.8 (m-Komponententestmodell)**.** Gegeben sei das Modell multipler Testmessungen für $i = 1, ..., n$ Personen und $j = 1, ..., m$ Testmessungen. Für $i = 1, ..., n$ seien

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

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

$$\mathbb{P}(\tau_1, ..., \tau_n, y_1, ..., y_n) := \prod_{i=1}^{n} \mathbb{P}(\tau_i, y_i) = \prod_{i=1}^{n} \mathbb{P}(y_i|\tau_i)\mathbb{P}(\tau_i) \tag{1.90}$$

das m-Komponententestmodell, wenn gilt dass

$$\mathbb{P}(\tau_1, y_1) = \cdots = \mathbb{P}(\tau_n, y_n). \tag{1.91}$$

•

Zum Verständnis dazu, wie Cronbach's α die inneren Konsistenz eines Tests misst, bietet es sich zunächst an, für das m-Komponententestmodell die Reliabilität im Sinne der Formulierung von Aussage (1) in Theorem 1.5 zu definieren.

Definition 1.9 (Reliabilität von m-Komponententestmodellen)**.** Gegeben sei ein m-Komponententestmodell mit Observed-Score-Summe y_i und True-Score-Summe τ_i für $i = 1, ..., n$ Personen oder ein m-Parallelkomponententestmodell mit Observed-Score-Summe y_i und True-Score τ_i für $i = 1, ..., n$. Dann ist die Reliabilität des Modells definiert als

$$\mathrm{R} := \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_i)} \text{ für ein beliebiges } 1 \leq i \leq n. \tag{1.92}$$

•

Cronbach's α

Vor dem Hintegrund des m-Komponentenmodells definieren wir *Cronbach's* α wie folgt.

Definition 1.10 (Cronbach's α). Gegeben sei ein m-Komponententestmodell. Dann heißt

$$\alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_{ij})}{\mathbb{V}(y_i)}\right) \tag{1.93}$$

Cronbach's α oder *Koeffizient* α.

•

Man beachte, dass in Definition 1.10 $\mathbb{V}(y_i)$ die Varianzen der Observed-Score-Summen für Personen $i = 1, ..., n$ und $\mathbb{V}(y_{ij})$ die Varianzen des Observed-Scores für Personen $i = 1, ..., n$ und Items $j = 1, ..., m$ bezeichnen.

INTUITIVE BEDEUTUNG DER FORMEL FÜR CRONBACHS ALPHA / GRENZ SZENARIEN BEZ. VIJ UND VI

Seine zentrale Bedeutung in der Klassischen Testtheorie bekommt Cronbach's α durch seinen Zusammenhang mit der Reliabilität eines m-Komponententests, welchen wir in unterem Theorem darstellen.

Theorem 1.8 (Cronbach's α und Reliabilität). *Gegeben sei ein m-Komponententestmodell mit Reliabilität R. Dann gilt für Cronbach's α, dass*

$$\alpha \le R \tag{1.94}$$

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

∘

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

$$y_j := y_{ij}, \tau_j := \tau_{ij}, y := \sum_{j=1}^m y_j \text{ und } \tau := \sum_{j=1}^m \tau_j, \tag{1.95}$$

sowie

$$\mathrm{R} := \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} \text{ und } \alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^m \mathbb{V}(y_j)}{\mathbb{V}(y)}\right). \tag{1.96}$$

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

$$\sum_{j=1}^m x_j = \frac{1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^m (x_j + x_k). \tag{1.97}$$

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

$$
\begin{aligned}
\frac{1}{3}\sum_{j=1}^{3}\sum_{k=j+1}^{4}(x_j + x_k) & = \frac{1}{3}\left(\sum_{k=1+1}^{4}(x_1 + x_k) + \sum_{k=2+1}^{4}(x_2 + x_k) + \sum_{k=3+1}^{4}(x_3 + x_k)\right) \\
& = \frac{1}{3}\left(\sum_{k=2}^{4}(x_1 + x_k) + \sum_{k=3}^{4}(x_2 + x_k) + \sum_{k=4}^{4}(x_3 + x_k)\right) \\
& = \frac{1}{3}\left((x_1 + x_2) + (x_1 + x_3) + (x_1 + x_4) + (x_2 + x_3) + (x_2 + x_4) + (x_3 + x_4)\right) \\
& = \frac{1}{3}\left((x_1 + x_1 + x_1) + (x_2 + x_2 + x_2) + (x_3 + x_3 + x_3) + (x_4 + x_4 + x_4)\right) \\
& = \frac{1}{3}\left(3x_1 + 3x_2 + 3x_3 + 3x_4\right) \\
& = x_1 + x_2 + x_3 + x_4 \\
& = \sum_{j=1}^{4} x_j
\end{aligned}
\tag{1.98}
$$

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

$$
\mathbb{V}(\tau_j - \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) - 2\mathbb{C}(\tau_j, \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) \geq 2\mathbb{C}(\tau_j, \tau_k) \tag{1.99}
$$

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

$$
\mathbb{V}(\tau_j - \tau_k) = \mathbb{V}(f_j(\tau_1) - f_k(\tau_1)) = \mathbb{V}(\tau_1 - \tau_1) = \mathbb{V}(0) = 0. \tag{1.100}
$$

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

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

$$
\begin{aligned}
\mathbb{V}(\tau) & = \mathbb{V}\left(\sum_{j=1}^{m}\tau_j\right) \\
& = \sum_{j=1}^{m}\mathbb{V}(\tau_j) + 2\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\mathbb{C}(\tau_j, \tau_k) \\
& = \frac{1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\left(\mathbb{V}(\tau_j) + \mathbb{V}(\tau_k)\right) + 2\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\mathbb{C}(\tau_j, \tau_k) \\
& \geq \frac{1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}2\mathbb{C}(\tau_j, \tau_k) + \sum_{j=1}^{m-1}\sum_{k=j+1}^{m}2\mathbb{C}(\tau_j, \tau_k) \\
& = \left(\frac{1}{m-1} + 1\right)\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\mathbb{C}(\tau_j, \tau_k) \\
& = \left(\frac{1}{m-1} + \frac{m-1}{m-1}\right)\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}2\mathbb{C}(\tau_j, \tau_k) \\
& = \frac{1+m-1}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}2\mathbb{C}(\tau_j, \tau_k) \\
& = \frac{m}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m}2\mathbb{C}(\tau_j, \tau_k)
\end{aligned}
\tag{1.101}
$$

Mit Aussage (5) von Theorem 1.3
und wiederum mit dem Theorem zur Varianz einer Linearkombination LINK gilt dann

$$
\begin{aligned}
\mathbb{V}(\tau) &\geq \frac{m}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j,\tau_k) \\
&= \frac{m}{m-1}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} 2\mathbb{C}(y_j,y_k) \\
&= \frac{m}{m-1}\left(\mathbb{V}\left(\sum_{j=1}^{m} y_j\right) - \sum_{j=1}^{m}\mathbb{V}\left(y_j\right)\right) \\
&= \frac{m}{m-1}\left(\mathbb{V}(y) - \sum_{j=1}^{m}\mathbb{V}\left(y_j\right)\right)
\end{aligned}
\tag{1.102}
$$

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

$$
\begin{aligned}
\mathrm{R} &= \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} \\
&\geq \frac{\frac{m}{m-1}\left(\mathbb{V}(y) - \sum_{j=1}^{m}\mathbb{V}\left(y_j\right)\right)}{\mathbb{V}(y)} \\
&= \frac{m}{m-1}\left(\frac{\mathbb{V}(y)}{\mathbb{V}(y)} - \frac{\sum_{j=1}^{m}\mathbb{V}\left(y_j\right)}{\mathbb{V}(y)}\right) \\
&= \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m}\mathbb{V}\left(y_j\right)}{\mathbb{V}(y)}\right) \\
&=: \alpha.
\end{aligned}
\tag{1.103}
$$

□

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

Eine Schätzung von Cronbach's α greift dementsprechend auf die Stichprobenvarianzen der Summenscores und die Stichprobenvarianzen der Itemscores zurück.

$\mathbb{V}(y_i)$ ist die Varianz der Observed-Score Summen für Personen $i = 1, ..., n$.

Es gilt $\mathbb{V}(y_1) = \cdots = \mathbb{V}(y_n)$.

Ein Schätzer für $\mathbb{V}(y_i)$ ist die Stichprobenvarianz der Observed-Score-Summe

$$
S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ mit } \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{1.104}
$$

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

Es gilt $\mathbb{V}(y_{1j}) = \cdots = \mathbb{V}(y_{nj})$ für $i = 1, ..., n$.

Ein Schätzer für $\mathbb{V}(y_{ij})$ ist die Stichprobenvarianz des Observed-Scores von Item j

$$
S_j^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_j\right)^2 \text{ mit } \bar{y}_j := \frac{1}{n}\sum_{i=1}^{n} y_{ij}. \tag{1.105}
$$

THEOREM UNVERZERRTER CRONBACHS ALPHA SCHTZER

Wir demonstrieren dies untenstehen anhand einer Simulation im Modell paralleler Testmessungen für $n := 30$ und $m := 21$ mit

$$\mathbb{P}(\tau_i) := N(1,1), \mathbb{P}(y_{ij}|\tau_i) := N(\tau_i, 4) \text{ für } i = 1, ..., n, j = 1, ..., m. \tag{1.106}$$

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

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

Referenzen

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

Fisher, R. A. (1925). Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *22*(5), 700–725. https://doi.org/10.1017/S0305004100009580

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 2* (2nd ed). Wiley.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores* (Nachdr. der Ausg. Reading, Mass. [u.a.], 1968). Information Age Publ.

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