Varianzzerlegung

- Die Varianz der beobachteten Testwerte x_v: setzt sich zusammen aus zerlegen wahrer Varianz und Fehlervarianz:

$$\sigma^2(x) = \sigma^2(\tau) + \sigma^2(\varepsilon)$$

Varianzzerlegung und Definition der Reliabilität

$$\sigma^2(x) = \sigma^2(\tau) + \sigma^2(\varepsilon)$$

- Die Reliabilität (Messgenauigkeit) eines Test ist definiert als der Anteil der beobachteten Varianz in den Testwerten, der auf Variation in den wahren Testwerten der Testpersonen zurückgeht:

$$\mathrm{Rel} = \frac{\sigma^2(\tau)}{\sigma^2(x)} = \frac{\sigma^2(x) - \sigma^2(\varepsilon)}{\sigma^2(x)}$$

Varianzzerlegung und Definition der Reliabilität

$$\sigma^2(x) = \sigma^2(\tau) + \sigma^2(\varepsilon)$$

- Für $\sigma^2(x) = \sigma^2(\tau)$:

$$\mathrm{Rel} = \frac{\sigma^2(\tau)}{\sigma^2(x)} = \frac{\sigma^2(\tau)}{\sigma^2(\tau)} = 1$$

„Der Test misst völlig fehlerfrei.“

- Für $\sigma^2(x) = \sigma^2(\varepsilon)$:

$$\sigma^2(\tau) = \sigma^2(x) - \sigma^2(\varepsilon) = 0 \rightarrow \mathrm{Rel} = \frac{0}{\sigma^2(x)} = 0$$

„Der Test misst gar nichts.“

Varianzzerlegung und Definition der Reliabilität

- Die Reliabilität eines auf Basis der KTT konstruierten Tests wird durch einen Reliabilitätskoeffizienten beschrieben.
- Dieser Koeffizient hat einen theoretischen Wertebereich von null bis eins,
- eins bedeutet hierbei eine „perfekte“, d.h. messfehlerfreie Messung,
- null bedeutet, dass die Varianz zwischen den Testwerten vollständig zufällig ist.

Grundlagen der Reliabilitätsschätzung

- Im Rahmen der Methoden der KTT existiert eine Vielzahl von Methoden zur Schätzung des Reliabilitätskoeffizienten.
- Im Folgenden werden zunächst die gemeinsamen Grundlagen dieser Schätzverfahren dargestellt.
- Es wird gezeigt, dass die Reliabilität eines Tests sich schätzen lässt als die Korrelation eines Tests mit sich selbst.

Grundlagen der Reliabilitätsschätzung

- Da die Messfehler ε in der KTT als vollkommen zufällig und untereinander unkorreliert angenommen werden, ist die Kovarianz zweier Tests x_p und x_q gleich der Kovarianz ihrer wahren Werte τ_p und τ_q:

$$\sigma\left(x_{p,}, x_q\right) = \sigma\left(\tau_p + \varepsilon_p, \tau_q + \varepsilon_q\right) =$$

$$\sigma\left(\tau_p, \tau_q\right) + \sigma\left(\tau_p, \varepsilon_q\right) + \sigma\left(\tau_q, \varepsilon_p\right) + \sigma\left(\varepsilon_q, \varepsilon_p\right) =$$

$$\sigma\left(\tau_p, \tau_q\right)$$

Grundlagen der Reliabilitätsschätzung

- Entsprechend lässt sich die Korrelation zweier Tests als der Anteil „wahrer Kovarianz“ an ihrer beobachteten Gesamtvarianz ausdrücken:

$$\rho(x_p, x_q) = \frac{\sigma(x_p, x_q)}{\sigma(x_p)\cdot\sigma(x_q)}$$

Korrelation der Testwerte x_p und x_q

$$\rho(x_p, x_q) = \frac{\sigma(\tau_p, \tau_q)}{\sigma(x_p)\cdot\sigma(x_q)}$$

Grundlagen der Reliabilitätsschätzung

- Im Fall zweier Messungen mit demselben Test (desselben „true scores“), ist die Korrelation der Messungen gleich dem Anteil der Varianz ihrer wahren Werte.

Grundlagen der Reliabilitätsschätzung

$$\rho(x_p, x_q) = \frac{\sigma(\tau_p, \tau_q)}{\sigma(x_p)\cdot\sigma(x_q)}$$

Bei zwei Messungen mit demselben Test x_p:

$$\sigma(x_p, x_p) = \sigma(\tau_p, \tau_p) = \sigma^2(\tau_p)$$

$$\sigma(x_p)\ \cdot\ \sigma(x_p) = \sigma^2(x_p)$$

$$\rho(x_p, x_p) = \frac{\sigma^2(\tau_p)}{\sigma^2(x_p)}$$

→ Die Korrelation eines Tests mit sich selbst entspricht seiner Reliabilität.

Grundlagen der Reliabilitätsschätzung

- Der als Korrelation zweier Messungen mit demselben Test bestimmte Reliabilitätskoeffizient kann Werte zwischen 0 und 1 annehmen. Mit Hundert multipliziert gibt er den prozentualen Anteil wahrer Varianz an der Testwertvarianz an (Determinationskoeffizient).
- Vorsicht: Im Unterschied zu Korrelations-koeffizienten wird der Reliabilitätskoeffizient nicht quadriert!

Grundlagen der Reliabilitätsschätzung: Reliabilität und Testlänge

- Aus den Annahmen der KTT lässt sich ableiten, dass die Verlängerung eines Tests durch Hinzunahme weiterer τ-äquivalenter Testteile zu einer Erhöhung der Reliabilität führt.
- Beispiel: Ein Test x_p und ein Test x_q erfassen dasselbe Merkmal und haben dieselben Anteile wahrer Varianz (*parallele Tests*).
- Beide Tests werden zu einem neuen Gesamttest zusammengefügt, d.h. die Testlänge verdoppelt sich.

Grundlagen der Reliabilitätsschätzung: Reliabilität und Testlänge

- Zwei parallele Tests x_p und x_q werden zu einem neuen Gesamttest zusammengefügt. Die Varianz des neuen Tests ergibt sich wie folgt:

$$\sigma^2\left(x_p + x_q\right) =$$

$$\sigma^2\left(x_p\right) + \sigma^2\left(x_q\right) + 2\sigma\left(x_p, x_q\right)$$

Grundlagen der Reliabilitätsschätzung: Reliabilität und Testlänge

- Die Kovarianz zweier Tests gleich der Kovarianz ihrer wahren Werte τ_p und τ_q. Wenn den Messungen x_p und x_q derselbe true score τ zugrunde liegt, ist die Kovarianz gleich der wahren Varianz.

$$\sigma\left(x_p, x_q\right) = \sigma^2\left(\tau\right)$$

- Der Varianz der beiden parallelen Tests liegt das gleiche Verhältnis von wahrer und Fehlervarianz zugrunde:

$$\sigma^2\left(x_p\right) = \sigma^2\left(x_q\right) = \sigma^2\left(\tau\right) + \sigma^2\left(\varepsilon\right)$$

Grundlagen der Reliabilitätsschätzung: Reliabilität und Testlänge

- Für die Zusammensetzung der Varianz des verlängerten Test $x_p + x_q$ ergibt sich also:

$$\sigma^2\left(x_1 + x_2\right) =$$
$$\sigma^2\left(x_1\right) + \sigma^2\left(x_2\right) + 2\sigma\left(x_1, x_2\right) =$$
$$\sigma^2\left(\tau\right) + \sigma^2\left(\varepsilon\right) + \sigma^2\left(\tau\right) + \sigma^2\left(\varepsilon\right) + 2\sigma^2\left(\tau\right) =$$
$$4\sigma^2\left(\tau\right) + 2\sigma^2\left(\varepsilon\right)$$

- → Die wahre Varianz vervierfacht sich, während sich die Fehlervarianz lediglich verdoppelt.

Grundlagen der Reliabilitätsschätzung: Reliabilität und Testlänge

- Der Zuwachs an Reliabilität bei Verdoppelung der Testlänge eines Tests xp lässt wie folgt ausdrücken:

$$\mathrm{Rel}\left(2 \cdot x_p\right) = \frac{2 \cdot \mathrm{Rel}\left(x_p\right)}{1 + \mathrm{Rel}\left(x_p\right)}$$

- Allgemeiner bei Verlängerung eines Tests um den Faktor k („Spearman-Brown-Formel zur Testverlängerung“):

$$\mathrm{Rel}\left(k \cdot x_p\right) = \frac{k \cdot \mathrm{Rel}\left(x_p\right)}{1 + \left(k-1\right) \cdot \mathrm{Rel}\left(x_p\right)}$$

Methoden zur Schätzung der Reliabilität

- Die zur Schätzung der Reliabilität nötigen „zwei Messungen mit demselben Test“ können auf unterschiedliche Weisen realisiert werden. Die verschiedenen in der Praxis eingesetzten Verfahren zur Reliabilitätsbestimmung gehen fast alle auf das Prinzip zurück, zwei oder mehrere Messungen, denen derselbe true score τ zugrunde liegt, miteinander zu korrelieren.

Methoden zur Schätzung der Reliabilität

Ermittlung des Reliabilitätskoeffizienten

- Retest-Reliabilität
- Paralleltest-Reliabilität
- Interne Konsistenz
  - Halbtest-Reliabilität („split-half“)
  - Alpha-Koeffizient (Cronbachs α)
  - Varianzanalytische Reliabilitätsschätzung

Retest-Reliabilität

- Die einfachste Art, die Korrelation eines Tests mit sich selbst zu bestimmen, ist, tatsächlich zwei Messungen x_{t1} und x_{t2} mit demselben Test vorzunehmen.

$$\rho(x_{t1}, x_{t2}) = \frac{\sigma(\tau_{t1}, \tau_{t2})}{\sigma(x_{t1}) \cdot \sigma(x_{t2})} = \frac{\sigma^2(\tau)}{\sigma^2(x)}$$

$$\widehat{\mathrm{Rel}} = r\left(x_{t1}, x_{t2}\right)$$

Retest-Reliabilität

- Es ist offensichtlich, dass diesem Vorgehen praktische Grenzen gesetzt sind.
- Vor allem bei Leistungstests können Erinnerungseffekte nicht ausgeschlossen werden.
- Bei Persönlichkeitsfragebögen ist die Ermittlung der Retest-Reliabilität ein verbreitetes Vorgehen.
- Die Retest-Reliabilität ist nur dann eine sinnvolle Schätzung der Messgenauigkeit, wenn das erfasste Merkmal als zeitlich stabil betrachtet wird!

Paralleltest-Reliabilität

- Zwei Tests x_p und x_q werden als „parallel" bezeichnet, wenn
- beiden Messungen derselbe true score t zugrunde liegt: $E(x_p) = E(x_q) = \tau$
- beide Tests gleiche Streuungen und gleiche Anteile von wahrer und Fehlervarianz aufweisen: $\sigma^2(x_p) = \sigma^2(x_q) = \sigma^2(\tau) + \sigma^2(\varepsilon)$.
- Parallele Tests sind typischerweise dann verfügbar, wenn für ein Testverfahren von Anfang an zwei parallele Versionen konstruiert wurden.

Paralleltest-Reliabilität

- Existieren zwei parallele Fassungen x_p und x_q eines Tests, kann die Reliabilität des Verfahrens geschätzt werden als die Korrelation zwischen beiden Testfassungen:

$$\rho\left(x_p, x_q\right) = \frac{\sigma\left(\tau_p, \tau_q\right)}{\sigma\left(x_p\right)\cdot\sigma\left(x_q\right)} = \frac{\sigma^2\left(\tau\right)}{\sigma^2\left(x\right)}$$

$$\widehat{Rel} = r\left(x_p, x_q\right)$$

Interne Konsistenz

- Bei der Bestimmung der internen Konsistenz werden zwei oder mehrere *Testteile* als τ-äquivalente Messungen betrachtet.
- Zur Schätzung der Reliabilität werden die Zusammenhänge zwischen den Testteilen herangezogen.
- Die interne Konsistenz kann mit verschiedenen Methoden ermittelt werden:
  - Halbtest-Reliabilität („split-half“)
  - Alpha-Koeffizient (Cronbachs α)
  - Konsistenzanalytische Reliabilitätsschätzung

Interne Konsistenz: Split-Half-Reliabilität

- Zur Bestimmung der Split-Half-Reliabilität wird der untersuchte Test in zwei gleich lange Hälften x_1 und x_2 geteilt.
- Im einfachsten Fall können die Items abwechselnd nach gerader und ungerader Itemposition ("odd-even") auf die Hälften aufgeteilt werden.
- Häufig werden die Itemkennwerte des Tests herangezogen, um möglichst parallele Testformen zu bilden.

Interne Konsistenz: Split-Half-Reliabilität

- Die Reliabilität wird durch die Korrelation der beiden Testhälften x_1 und x_2 geschätzt.
- Diese Schätzung wird nach der Spearman-Brown-Formel auf die doppelte Testlänge aufgewertet.

$$\widehat{\mathrm{Rel}} = \frac{2 \cdot r(x_1, x_2)}{1 + r(x_1, x_2)}$$

Interne Konsistenz: Cronbachs α

- Bei der Bestimmung von Cronbachs α wird der Test nicht nur in zwei, sondern in k Testteile zerlegt und die Summe der Varianzen der Testteile zur Varianz des Gesamttests in Relation gesetzt:

$$\widehat{\mathrm{Rel}} = \alpha = \frac{k}{k-1} \cdot \left(1 - \frac{\sum_{j=1}^{k} \sigma^2(x_k)}{\sigma^2(x)} \right)$$

k = Anzahl der Testteile; $\sigma^2(x_j)$ = Varianz des Testteils j; $\sigma^2(x)$ = Varianz des Gesamttests

Interne Konsistenz: Cronbachs α

- Cronbachs α beruht auf dem Gedanken, dass jede Kovarianz zwischen beliebigen Testteilen als wahre Varianz $\sigma^2(\tau)$ betrachtet werden kann.
- Je stärker die Testteile positiv kovariieren, desto mehr nähert sich Alpha dem Wert eins an, da die Varianz des Gesamttests größer ist als die Summe der Varianzen der Testteile.
- Wenn die Testteile keine Kovariation aufweisen, wird Alpha null, da die Varianz des Gesamttests dann gleich der Summe der Varianzen der Teiltests ist.

Interne Konsistenz: Cronbachs α

- Zur Bestimmung von Cronbachs α wird der Test i.d.R. in so viele Teile zerlegt, wie er Items enthält, d.h. jedes Item wird als Testteil behandelt.
- Cronbachs α ist der für auf der KTT basierende Tests am häufigsten verwendete Reliabilitätskoeffizient.

Interne Konsistenz: Varianzanalytische Reliabilitätsbestimmung

- Bei der varianzanalytischen Reliabilitätsbestimmung wird eine Zerlegung der Antwortvarianz Q_{tot} vorgenommen in denjenigen Varianzanteil Q_{Items}, welcher aus den unterschiedlich hohen Itemanforderungen (Itemschwierigkeiten) resultiert, und denjenigen Anteil $Q_{Personen}$, der auf Unterschiede zwischen den Personen zurückgeführt werden kann.
- Die Varianz $Q_{Personen}$ zwischen den Personen stellt im Sinne der Klassischen Testtheorie die interessierende Merkmalsvarianz dar.

Interne Konsistenz: Varianzanalytische Reliabilitätsbestimmung

- Der als "Rest" verbleibende, weder auf Unterschiede zwischen den Items noch auf Unterschiede zwischen den Personen rückführbare Varianzanteil an der Testwertvarianz kann als Schätzer für den Fehlervarianzanteil herangezogen werden.
- Die wahre Varianz wird geschätzt über den Anteil der Fehlervarianz an der Testpersonen-Varianz. Je kleiner Q_{Rest}, desto größer die Reliabilität:

Interne Konsistenz: Varianzanalytische Reliabilitätsbestimmung

$$\widehat{\mathrm{Rel}} = 1 - \frac{\left(Q_{Tot} - Q_{Personen} - Q_{Items}\right)/df_{Rest}}{Q_{Personen}/df_{Personen}} =$$

$$1 - \frac{Q_{Rest}/df_{Rest}}{Q_{Personen}/df_{Personen}}$$

Q_{Tot} = Quadratsumme der Testwerte
$Q_{Personen}$ = Personen-Quadratsumme;
$df_{Personen}$ = Personenfreiheitsgrade = $N - 1$
Q_{Items} = Itemquadratsumme
Q_{Rest} = Fehlerquadratsumme;
df_{Rest} = Fehlerfreiheitsgrade = $(N - 1)\cdot(k - 1)$

Interne Konsistenz: Varianzanalytische Reliabilitätsbestimmung

- Die Varianzanalytische Reliabilitätsbestimmung wird selten zur Bestimmung der internen Konsistenz eingesetzt.

Interne Konsistenz

- Die Schätzung der Reliabilität durch die interne Konsistenz eines Tests hat mehrere praktische Vorteile:
  - Es muss nur eine Messung erfolgen (→ weniger Aufwand, Erinnerungseinflüsse sind kein Problem).
  - Es müssen keine Parallelformen eines Tests konstruiert werden.
  - Das erfasste Merkmal muss nicht zeitlich stabil sein.
- Die interne Konsistenz, insbesondere Cronbachs α, gehört daher zu den am meisten berichteten Reliabilitätsschätzungen.