Psychologische Testtheorie

Sitzung 11

Reliabilität II

Wahr-oder-Falsch Aussagen zur letzten Sitzung

1. Die Reliabilität eines Items entspricht dem Anteil der systematischen Varianz an der Gesamtvarianz der Itemantwort.
2. Wenn die Varianz der Fehlervariable steigt, nimmt auch die Reliabilität zu.
3. Der konkrete Wertebereich der Reliabilität hängt von der Einheit der Items ab.
4. Die Schätzung der Reliabilität eines Items ist unabhängig vom geltenden Messmodell.
5. Im parallelen Modell weisen alle Items des Tests dieselbe Reliabilität auf.
6. Die Reliabilität eines Items i entspricht im essentiell parallelen Modell der Korrelation dieses Items mit einem beliebigen anderen Item j des Tests.

Überblick

Sitzung	Datum	Thema	Themenblock
1	13.10.25	Einführung	Begriffe, Modellierung von Antwortverhalten durch Zufallsvariablen & mathematische Grundlagen der Testtheorie
2	20.10.25	Wahrscheinlichkeitstheoret. Grundlagen	
3	27.10.25	Testtheoretische Modelle I	Testtheoretische Modelle
4	03.11.25	Testtheoretische Modelle II	
5	10.11.25	Testtheoretische Modelle III	
6	17.11.25	Skalierung I	Gütekriterien psychologischer Tests
7	24.11.25	Skalierung II	
8	01.12.25	Faktorenanalyse I	
9	08.12.25	Faktorenanalyse II	
10	15.12.25	Reliabilität I	
11	22.12.25	Sitzung entfällt wegen Weihnachten!	
	30.12.25	Offizielle Winterpause	
	06.01.26		
12	12.01.26	Reliabilität II	

➛ In der heutigen Vorlesung beschäftigen wir uns weiter mit der Reliabilität – diesmal auf Ebene des gesamten Tests.

2.8. Schätzung der Reliabilität

Übersicht: Reliabilität einzelner Items

- Paralleles Modell: $REL(X_i) = COR\left(X_i, X_j\right)$
- Essentiell paralleles Modell: $REL(X_i) = COR\left(X_i, X_j\right)$
- τ-äquivalentes Modell: $REL(X_i) = \frac{COV\left(X_i, X_j\right)}{VAR(X_i)}$
- Essentiell τ- äquivalentes Modell: $REL(X_i) = \frac{COV\left(X_i, X_j\right)}{VAR(X_i)}$
- τ-kongenerisches Modell: $REL(X_i) = \beta_{zi}^2$
- Mehrdimensionales τ-kongenerisches Modell: $REL(X_i) = \sum_{l=1}^{q} \beta_{zil}^2$

Paralleles und essentiell paralleles Modell I

- Die Reliabilität eines Items i entspricht im **parallelen und essentiell parallelen Modell** der Korrelation dieses Items mit einem beliebigen anderen Item j:

$$REL(X_i) = COR(X_i, X_j)$$

- Da das Item j beliebig ist, kommen zunächst verschiedene Korrelationen als Schätzwerte für die Reliabilität eines Items i in Frage
- Z.B. könnte als Schätzwert für die Reliabilität von Item 1 die beobachtete Korrelation zwischen Item 1 und Item 2 oder die beobachtete Korrelation zwischen Items 1 und Item 3 verwendet werden
- Da jedoch die Reliabilitäten (und die Korrelationen) aller Items im parallelen und essentiell parallelen Modell gleich sind, können wir diese auch kombiniert durch den Mittelwert aller beobachtete Korrelationen schätzen

Paralleles und essentiell paralleles Modell II

- Beispiel: 4 Items, die einem (essentiell) parallelen Modell folgen
- Korrelationstabelle der 4 TAP-Items:

	almd1	almd2	almd3	almd4
almd1	1.00	**0.74**	**0.63**	**0.62**
almd2	0.74	1.00	**0.85**	**0.70**
almd3	0.63	0.85	1.00	**0.78**
almd4	0.62	0.70	0.78	1.00

- Wir haben also (0.74 + 0.63 + 0.62 + 0.85 + 0.70 + 0.78) / 6 = 0.72 als Schätzwert für die Reliabilität für jedes einzelne Item

Paralleles und essentiell paralleles Modell III

- Es ergeben sich somit als Schätzwerte für die Reliabilitäten der Items:
  - Item 1: $rel(x_1) = 0.72$
  - Item 2: $rel(x_2) = 0.72$
  - Item 3: $rel(x_3) = 0.72$
  - Item 4: $rel(x_4) = 0.72$

- Wir verwenden hier die Notation $rel(x_i)$ statt $REL(X_i)$ um zu verdeutlichen, dass es sich hier um <u>Schätzwerte</u> der unbekannten tatsächlichen Reliabilitäten $REL(X_i)$ handelt

τ-äquivalentes und essentiell τ-äquivalentes Modell I

- Die Reliabilität eines Items i entspricht im **$\boldsymbol{\tau}$-äquivalenten und essentiell $\boldsymbol{\tau}$-äquivalenten Modell** der Kovarianz dieses Items mit einem beliebigen anderen Item j geteilt durch seine Varianz:

$$REL(X_i) = \frac{COV(X_i, X_j)}{VAR(X_i)}$$

- Die Varianzen $VAR(X_i)$ der Items können wir jeweils durch die beobachteten Varianzen der Items schätzen
- Da die Kovarianzen aller Itempaare im τ-äquivalenten und essentiell τ-äquivalenten Modell gleich sind (siehe Folgerungen aus den Modellannahmen), können wir diese kombiniert durch den Mittelwert aller beobachteten Kovarianzen schätzen

τ-äquivalentes und essentiell τ-äquivalentes Modell II

- Beispiel: 4 Items, die einem (essentiell) τ-äquivalenten Modell folgen
- Kovarianztabelle der 4 TAP-Items:

	almd1	almd2	almd3	almd4
almd1	5767.57	3538.36	2894.04	3095.98
almd2	3538.36	3969.41	3236.35	2898.61
almd3	2894.04	3236.35	3664.66	3115.13
almd4	3095.98	2898.61	3115.13	4371.38

- Als Schätzwert für die unbekannte Varianz für Item 1 ergibt sich 5767.57, für Item 2 3969.41, für Item 3 3664.66 und für Item 4 schließlich 4371.38
- Als Schätzwert für die unbekannte Kovarianz zweier beliebiger Items: (3538.36 + 2894.04 + 3095.98 + 3236.35 + 2898.61 + 3115.13) / 6 = 3129.74

τ-äquivalentes und essentiell τ-äquivalentes Modell III

- Es ergeben sich somit als Schätzwerte für die Reliabilitäten der Items:

	almd1	almd2	almd3	almd4
almd1	5767.57	3538.36	2894.04	3095.98
almd2	3538.36	3969.41	3236.35	2898.61
almd3	2894.04	3236.35	3664.66	3115.13
almd4	3095.98	2898.61	3115.13	4371.38

  - Item 1: $rel(x_1) = \frac{3129.74}{5767.57} = 0.54$
  - Item 2: $rel(x_2) = \frac{3129.74}{3969.41} = 0.79$
  - Item 3: $rel(x_3) = \frac{3129.74}{3664.66} = 0.85$
  - Item 4: $rel(x_4) = \frac{3129.74}{4371.38} = 0.72$

τ-kongenerisches Modell I

- Die Reliabilität eines Items i entspricht im **$\boldsymbol{\tau}$-kongenerischen Modell** dem quadrierten standardisierten Steigungsparameter dieses Items:

$$REL(X_i) = \beta_{zi1}^2$$

- Die geschätzten standardisierten Steigungsparameter können wir dem R-Output mit den Parameterschätzungen entnehmen
- Diese können wir dann quadrieren, um Schätzwerte für die Reliabilitäten der Items zu erhalten

τ-kongenerisches Modell II

- Beispiel: 4 Items, die einem τ-kongenerisches Modell an
- R Output der CFA der 4 TAP-Items

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  f1 =~
    almd1            56.526    5.649   10.006    0.000   56.526    0.747
    almd2            57.754    4.197   13.759    0.000   57.754    0.920
    almd3            55.512    4.032   13.767    0.000   55.512    0.920
    almd4            52.933    4.759   11.124    0.000   52.933    0.804
```

- Der Schätzwert für die Reliabilität des ersten Items wäre hier z.B. $rel(x_1) = 0.747^2 \approx 0.56$

Mehrdimensionales τ-kongenerisches Modell I

- Die Reliabilität eines Items i entspricht im **mehrdimensionalen $\boldsymbol{\tau}$-kongenerischen Modell** der Kommunalität dieses Items:

$$REL(X_i) = \sum_{l=1}^{q} \beta_{zil}^2$$

- Die Kommunalitäten – und somit die Reliabilitäten – der Items können im Rahmen der Exploratorischen Faktorenanalyse geschätzt werden
- Dies gilt natürlich auch für das eindimensionale τ-kongenerische Modell, da dieses einen Spezialfall des mehrdimensionalen Modells darstellt

- Beispiel: Items des NEO-FFI

```
Factor Analysis using method =  ml
Call: fa(r = NEO, nfactors = 2, rotate = "none", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
     ML1   ML2   h2   u2 com
n7   0.51 -0.06 0.27 0.73 1.0
n17  0.49  0.11 0.25 0.75 1.1
n27  0.26  0.02 0.07 0.93 1.0
n32  0.43  0.35 0.30 0.70 1.9
n37  0.83 -0.21 0.73 0.27 1.1
n42  0.65 -0.13 0.44 0.56 1.1
n47  0.13  0.44 0.21 0.79 1.2
n52  0.59  0.40 0.51 0.49 1.8
```

→ $0.51^2 + (-0.06)^2 \approx 0.27$

Achtung: Manuelle Berechnung der Kommunalität nur in der Anfangslösung vor der Rotation möglich (hier „rotate = none“)

3. Reliabilität des Itemmittelwerts

Reliabilität Itemmittelwert und Itemsumme I

- Bisher haben wir die Reliabilität einzelner Items eines Tests betrachtet
- Was können wir nun unter der Reliabilität eines gesamten Tests mit k Items verstehen?
- Wenn von der „Reliabilität eines Tests“ die Rede ist, ist meistens die Reliabilität des **Itemmittelwerts** $\boldsymbol{\bar{X} = \frac{1}{k}\sum_{i=1}^{k} X_i}$ oder die Reliabilität der **Itemsumme** $\boldsymbol{\sum_{i=1}^{k} X_i}$ gemeint

Reliabilität Itemmittelwert und Itemsumme II

- Die **Reliabilität des Itemmittelwerts** ist definiert als die Varianz des Mittelwerts der wahren Werte der Items geteilt durch die Varianz des Mittelwerts der Items:

$$REL(\bar{X}) = \frac{VAR\left(\frac{1}{k}\sum_{i=1}^{k} \tau_i\right)}{VAR\left(\frac{1}{k}\sum_{i=1}^{k} X_i\right)}$$

- Die **Reliabilität der Itemsumme** ist definiert als die Varianz der Summe der wahren Werte der Items geteilt durch die Varianz der Summe der Items:

$$REL\left(\sum_{i=1}^{k} X_i\right) = \frac{VAR\left(\sum_{i=1}^{k} \tau_i\right)}{VAR\left(\sum_{i=1}^{k} X_i\right)}$$

Reliabilität Itemmittelwert und Itemsumme III

- Die beiden Reliabilitäten sind als Maße für die Genauigkeit eines Tests austauschbar, da

$$REL(\bar{X}) = \frac{VAR\left(\frac{1}{k}\sum_{i=1}^{k}\tau_i\right)}{VAR\left(\frac{1}{k}\sum_{i=1}^{k}X_i\right)} = \frac{\frac{1}{k^2}VAR\left(\sum_{i=1}^{k}\tau_i\right)}{\frac{1}{k^2}VAR\left(\sum_{i=1}^{k}X_i\right)} = \frac{VAR\left(\sum_{i=1}^{k}\tau_i\right)}{VAR\left(\sum_{i=1}^{k}X_i\right)} = REL\left(\sum_{i=1}^{k}X_i\right)$$

- Wir werden uns daher für den Rest der Vorlesung auf die Reliabilität des Itemmittelwerts beschränken
- Ob und inwiefern die Reliabilität des Itemmittelwerts bzw. der Itemsumme ein geeignetes Maß für die Genauigkeit eines psychologischen Tests ist, werden wir in der letzten Vorlesung zur **Einzelfalldiagnostik** noch einmal kritisch diskutieren
- Für die Beurteilung der Höhe der Reliabilität des Itemmittelwerts gibt es keine einheitlichen Regeln. Oft werden Werte ab .70 als „ausreichend“ und ab .80 als „gut“ interpretiert. Gerade in der Einzelfalldiagnostik sind jedoch mitunter noch höhere Werte erforderlich (siehe Sitzungen #14).

Schätzung der Reliabilität Itemmittelwert

- Die Reliabilität des Itemmittelwerts ist eine unbekannte Größe:

$$REL(\bar{X}) = \frac{VAR\left(\frac{1}{k}\sum_{i=1}^{k}\tau_i\right)}{VAR\left(\frac{1}{k}\sum_{i=1}^{k}X_i\right)}$$

- Wir müssen sie daher mithilfe der Daten aus einer Stichprobe schätzen
- Hier haben wir das gleiche Problem wie bei der Reliabilität der einzelnen Items:
  - Die wahren Werte τ_i und auch ihr Mittelwert $\frac{1}{k}\sum_{i=1}^{k}\tau_i$ können nicht beobachtet und daher im Zähler $VAR\left(\frac{1}{k}\sum_{i=1}^{k}\tau_i\right)$ nicht geschätzt werden
  - Lösung wie bei der Reliabilität der Einzelitems: In Abhängigkeit von dem jeweiligen testtheoretischen Modell, das für die Items gilt, werden wir die Formel für die Reliabilität des Itemmittelwerts jeweils so umstellen, dass nur noch beobachtbare Größen darin vorkommen

3.1. Paralleles und essentiell paralleles Modell

Spearman-Brown-Formel I

- Für das parallele und essentiell parallele Modell ergibt sich für die Reliabilität des Itemmittelwerts die sogenannte **Spearman-Brown-Formel**:

$$REL(\bar{X}) = \frac{k^2}{k(k-1) + \sum_{i=1}^{k} \frac{1}{REL(X_i)}}$$

- Der Beweis hierfür ist aufwendig (siehe Kapitel 1 des Zusatzmaterials in Moodle)
- Mithilfe der Formel kann bei bekannten Reliabilitäten $REL(X_i)$ der einzelnen Items des Tests die Reliabilität des Itemmittelwerts berechnet werden

Spearman-Brown-Formel II

- Da im parallelen und essentiell parallelen Modell die Reliabilitäten der einzelnen Items für alle Items gleich sind, kann die Formel weiter vereinfacht werden:

$$REL(\bar{X}) = \frac{k^2}{k(k-1) + \sum_{i=1}^{k} \frac{1}{REL(X_i)}} = \frac{k^2}{k(k-1) + \frac{k}{REL(X_i)}} = \frac{k}{(k-1) + \frac{1}{REL(X_i)}}$$

- Anmerkung: Im parallelen und essentiell parallelen Modell erhöht jedes Item die Reliabilität des Itemmittelwerts (siehe Kapitel 2 des Zusatzmaterial in Moodle)

Beispiel

- In unserem vorherigen Beispiel (Folie #8) zum parallelen und essentiell parallelen Modell hatten sich als Schätzwerte für die Reliabilitäten der einzelnen Items folgende Werte ergeben:
  - Item 1: $rel(x_1) = 0.72$
  - Item 2: $rel(x_2) = 0.72$
  - Item 3: $rel(x_3) = 0.72$
  - Item 4: $rel(x_4) = 0.72$

- Mithilfe dieser Schätzwerte und der Spearman-Brown-Formel können wir nun einen <u>Schätzwert</u> für die Reliabilität des Itemmittelwerts bestimmen:

$$rel(\bar{x}) = \frac{k}{(k-1) + \frac{1}{rel(x_i)}} = \frac{4}{(4-1) + \frac{1}{0.72}} \approx 0.91$$

Cronbachs Alpha

- Die Spearman-Brown-Formel kann in eine alternative Form überführt werden:

$$REL(\bar{X}) = \frac{k}{k-1} \cdot \left(1 - \frac{\sum_{i=1}^{k} VAR(X_i)}{VAR\left(\sum_{i=1}^{k} X_i\right)}\right)$$

- Der Beweis hierfür ist wieder etwas aufwendig (siehe Kapitel 3 des Zusatzmaterial in Moodle)
- Diese alternative Formel wird als **Cronbachs Alpha** bezeichnet
- Sie ist einfacher zu schätzen, da die Reliabilitäten der einzelnen Items nicht geschätzt werden müssen.

$$\alpha = \frac{k}{k-1} \cdot \left(1 - \frac{\sum_{i=1}^{k} var(X_i)}{var\left(\sum_{i=1}^{k} X_i\right)}\right)$$

3.2. τ-äquivalentes und essentiell τ-äquivalentes Modell

Spearman-Brown-Formel I

- Auch für das τ-äquivalente und das essentiell τ-äquivalente Modell gilt für die Reliabilität des Itemmittelwerts die Spearman-Brown-Formel:

$$REL(\bar{X}) = \frac{k^2}{k(k-1) + \sum_{i=1}^{k} \frac{1}{REL(X_i)}}$$

- Da im τ-äquivalenten und essentiell τ-äquivalenten Modell die Reliabilitäten der einzelnen Items <u>nicht</u> für alle Items gleich sind, kann die Formel hier jedoch nicht weiter vereinfacht werden

Spearman-Brown-Formel II

- Anmerkung: Ob ein einzelnes Item des Tests die Reliabilität des Itemmittelwerts senkt oder erhöht, hängt von seiner Reliabilität ab
- Man kann mithilfe von Statistikprogrammen ermitteln, welche Items die geschätzte Reliabilität des Mittelwerts erhöhen
- **Items, die die Reliabilität des Mittelwerts senken, können entfernt werden, falls (!!) die Inhaltsvalidität dadurch nicht gefährdet wird**
  - Reduzieren Items die Reliabilität, sind aber für die Inhaltsvalidität essentiell, muss über eine Umformulierung der Items nachgedacht werden
  - Nach dem Entfernen eines Items aus einem Test muss eine Prüfung der Inhaltsvalidität des Tests erfolgen und ggf. die Konstruktbezeichnung angepasst werden

Beispiel

- In unserem vorherigen Beispiel (Folie #11) zum τ-äquivalenten und essentiell τ-äquivalenten Modell hatten sich als Schätzwerte für die Reliabilitäten der einzelnen Items folgende Werte ergeben:
  - Item 1: $rel(x_1) = 0.54$
  - Item 2: $rel(x_2) = 0.79$
  - Item 3: $rel(x_3) = 0.85$
  - Item 4: $rel(x_4) = 0.72$

- Mithilfe dieser Schätzwerte und der Spearman-Brown-Formel können wir nun einen <u>Schätzwert</u> für die Reliabilität des Itemmittelwerts bestimmen:

$$rel(\bar{x}) = \frac{k^2}{k(k-1) + \sum_{i=1}^{k} \frac{1}{rel(x_i)}} = \frac{4^2}{4(4-1) + \frac{1}{0.54} + \frac{1}{0.79} + \frac{1}{0.85} + \frac{1}{0.72}} \approx 0.90$$

Cronbachs Alpha

- Auch im τ-äquivalenten und essentiell τ-äquivalenten Modell kann alternativ zur Spearman-Brown-Formel wieder Cronbachs Alpha verwendet werden:

$$REL(\bar{X}) = \frac{k}{k-1} \cdot \left(1 - \frac{\sum_{i=1}^{k} VAR(X_i)}{VAR\left(\sum_{i=1}^{k} X_i\right)}\right) = \alpha$$

3.3. τ- kongenerisches Modell

McDonalds Omega I

- Im τ-kongenerischen Modell ergibt sich für die Reliabilität des Itemmittelwerts folgende Umformung:

$$REL(\bar{X}) = \frac{VAR\left(\frac{1}{k}\sum_{i=1}^{k}\tau_i\right)}{VAR\left(\frac{1}{k}\sum_{i=1}^{k}X_i\right)} = \frac{VAR\left(\sum_{i=1}^{k}\tau_i\right)}{VAR\left(\sum_{i=1}^{k}X_i\right)} = \frac{VAR\left(\sum_{i=1}^{k}(\sigma_i + \beta_i \cdot \theta)\right)}{VAR\left(\sum_{i=1}^{k}X_i\right)}$$

$$= \frac{VAR\left(\sum_{i=1}^{k}\sigma_i + \sum_{i=1}^{k}\beta_i \cdot \theta\right)}{VAR\left(\sum_{i=1}^{k}X_i\right)} = \frac{VAR\left(\sum_{i=1}^{k}\beta_i \cdot \theta\right)}{VAR\left(\sum_{i=1}^{k}X_i\right)} = \frac{VAR\left(\theta \cdot \sum_{i=1}^{k}\beta_i\right)}{VAR\left(\sum_{i=1}^{k}X_i\right)}$$

$$= \frac{\left(\sum_{i=1}^{k}\beta_i\right)^2 VAR(\theta)}{VAR\left(\sum_{i=1}^{k}X_i\right)} = \boxed{\frac{\left(\sum_{i=1}^{k}\beta_i\right)^2}{VAR\left(\sum_{i=1}^{k}X_i\right)}}$$

- Der letzte Bruch wird **McDonalds Omega** genannt

McDonalds Omega II

- Für das τ-kongenerische Modell ergibt sich die Reliabilität des Itemmittelwerts durch **McDonalds Omega**:

$$\omega = \frac{\left(\sum_{i=1}^{k} \beta_i\right)^2}{VAR\left(\sum_{i=1}^{k} X_i\right)}$$

- Hinweis: McDonalds Omega stellt auch in den strengeren Testmodellen eine angemessene Reliabilitätsschätzung dar

Cronbachs Alpha

- Auch im τ-kongenerischen Modell kann Cronbachs Alpha verwendet werden
- Anders als in den bisher besprochenen Modellen entspricht Cronbachs Alpha hier jedoch nicht genau der Reliabilität, sondern stellt lediglich **eine untere Schranke** der Reliabilität dar:

$$\alpha \leq \omega = REL(\bar{X})$$

- Der Beweis hierfür ist sehr schwierig
- Achtung: Auch wenn die Reliabilität (d.h., McDonalds Omega) immer größer als oder gleich groß wie Cronbachs Alpha ist, gilt dies nicht zwangsläufig für deren Schätzwerte!
- Obwohl Cronbachs Alpha nicht exakt der Reliabilität entspricht, kann es immer noch sinnvoll interpretiert werden: Falls sich ein hohes Cronbachs Alpha für einen Test ergibt, kann man vermuten, dass die tatsächliche Reliabilität noch höher ist
- In der Praxis gibt es Argumente für und gegen die Verwendung von Cronbachs Alpha im Fall τ-kongenerischer Modelle

Beispiel: R Output

- Aufgrund der potenziellen Ungenauigkeit der Schätzwerte für Cronbachs Alpha und McDonalds Omega sollte stets das **Konfidenzintervall** berücksichtigt werden
- Konfidenzintervalle für $\boldsymbol{\alpha}$ und $\boldsymbol{\omega}$ im TAP-Beispiel:

```
$est
[1] 0.9050416

$se
[1] 0.01606122

$ci.lower
[1] 0.8666673

$ci.upper
[1] 0.931773

$conf.level
[1] 0.95

$type
[1] "alpha"

$interval.type
[1] "bca bootstrap"
```

Punktschätzwert

Untere Grenze des Konfidenzintervalls

Obere Grenze des Konfidenzintervalls

Schätzer

```
$est
[1] 0.9046532

$se
[1] 0.01625434

$ci.lower
[1] 0.8645187

$ci.upper
[1] 0.9295699

$conf.level
[1] 0.95

$type
[1] "omega"

$interval.type
[1] "bca bootstrap"
```

3.4. Mehrdimensionales τ- kongenerisches Modell

Mehrdimensionales τ- kongenerisches Modell

- Auch im mehrdimensionalen τ-kongenerischen Modell lassen sich Formeln für die Reliabilität des Itemmittelwerts herleiten
- **Im Fall einer Einfachstruktur** ist es jedoch sinnvoller, die Items anhand dieser Struktur in mehrere eindimensionale τ-kongenerische Tests aufzuteilen und für jede Itemgruppe getrennt die Reliabilität des jeweiligen Mittelwertes auf Basis des eindimensionalen τ-kongenerischen Modells zu schätzen
- Da wir mehrere latente Variablen haben, sollten wir für jede auch eine eigene Reliabilität bestimmen – die Reliabilität des Gesamtmittelwerts aller Items stellt hier keine interessante Größe dar
- **Dies ist natürlich nicht möglich, wenn keine Einfachstruktur vorliegt!**

3.5. Weitere Methoden der Reliabilitätsschätzung

Interne Konsistenz Methode

- Die bisher besprochenen Methoden zur Schätzung der Reliabilität des Itemmittelwerts - abhängig vom jeweils geltenden Modell durch Spearman-Brown, Cronbachs Alpha oder McDonalds Omega - werden auch als Methoden der **internen / inneren Konsistenz** bezeichnet
- Da sie sich direkt aus den Modellen ergeben, die den Tests jeweils zugrunde liegen, sind sie in der Regel die empfohlenen Methoden zur Schätzung der Reliabilität eines Tests
- Es gibt noch weitere Methoden, die mittlerweile jedoch kaum noch verwendet werden und in der Regel nicht zu empfehlen sind. Aufgrund ihrer historischen Bedeutung besprechen wir sie dennoch.

Split-Half-Methode I

- Bei der **Split-Half-Methode** werden die Items eines Tests in zwei Hälften geteilt
- Der Mittelwert jeder Testhälfte wird als einzelnes „Item“ betrachtet
- Es wird die Annahme getroffen, dass die Mittelwerte der beiden Testhälften - als zwei neue „Items“ betrachtet - dem parallelen oder essentiell parallelen Modell genügen (die ursprünglichen Items des Tests müssen nicht unbedingt einem dieser beiden Modell folgen)
- Wir bezeichnen den Mittelwert der Items der ersten Testhälfte mit $\bar{X}_1$ und den Mittelwert der Items der zweiten Testhälfte mit $\bar{X}_2$
- Der Gesamtmittelwert aller Einzelitems ist dann der Mittelwert dieser beiden „Items“:
  - $\bar{X} = \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$

Split-Half-Methode II

- Falls die Mittelwerte der beiden Testhälften tatsächlich dem parallelen oder essentiell parallelen Modell folgen, können wir zur Schätzung der Reliabilität des Gesamtmittelwerts die vereinfachte Spearman-Brown-Formel für zwei parallele bzw. essentiell parallele Items anwenden:

$$REL(\bar{X}) = \frac{2}{(2-1) + \frac{1}{REL(\bar{X}_i)}}$$

Split-Half-Methode III

- Da wir annehmen, dass die Mittelwerte der Testhälften **parallele oder essentiell parallele „Items“** sind, entspricht die Reliabilität des Mittelwerts einer Testhälfte $REL(\bar{X}_i)$ (also die Reliabilität eines einzelnen „Items“) der Korrelation $COR(\bar{X}_1, \bar{X}_2)$ der Mittelwerte der beiden Testhälften und es ergibt sich folgende Vereinfachung der Formel auf der letzten Folie:

$$REL(\bar{X}) = \frac{2}{(2-1) + \frac{1}{REL(\bar{X}_i)}} = \frac{2}{1 + \frac{1}{COR(\bar{X}_1, \bar{X}_2)}} = \frac{2 \cdot COR(\bar{X}_1, \bar{X}_2)}{COR(\bar{X}_1, \bar{X}_2) + 1}$$

- Der große Nachteil dieser Methode besteht darin, dass wir die Parallelität der Mittelwerte der beiden Testhälften in einem wichtigen Punkt <u>nicht überprüfen können</u>: Wir können die dritte Folgerung (gleiche Kovarianzen aller Items) nicht testen, da nur eine Kovarianz zur Verfügung steht (es gibt ja nur zwei „Items“)

Paralleltest-Methode

- Ausgangssituation bei der **Paralleltest-Methode**: Es liegen Ergebnisse aus zwei unterschiedlichen Tests zur Erfassung der gleichen latenten Variable (i.d.R. zwei speziell entwickelte Parallelformen) vor (wie diese berechnet wurden, ist egal - es müssen nicht unbedingt die Itemmittelwerte sein) und die gleichen Personen haben beide Tests bearbeitet!
- Diese beiden Testergebnisse werden als „Items“ X_1 und X_2 aufgefasst
- Es wird die Annahme getroffen, dass die Testergebnisse X_1 und X_2 der beiden Tests dem parallelen oder essentiell parallelen Modell folgen (die ursprünglichen Items der beiden Tests müssen nicht unbedingt einem dieser Modelle folgen)
- Zur Schätzung der Reliabilität eines einzelnen „Items“ bzw. Tests kann dann wegen der Parallelität die Korrelation eines “Items“ bzw. Tests mit dem anderen “Item“ bzw. Test herangezogen werden: $REL(X_i) = COR(X_1, X_2)$
- Der große Nachteil dieser Methode besteht wieder darin, dass die dritte Folgerung aus der Parallelität (gleiche Kovarianzen aller Items) nicht überprüft werden kann

Retest-Methode

- Ausgangssituation bei der **Retest-Methode**: Ergebnisse des gleichen Tests zu zwei unterschiedlichen Zeitpunkten liegen vor (wie diese berechnet wurden, ist egal - es müssen nicht unbedingt die Itemmittelwerte sein) und die gleichen Personen haben den Test zweimal bearbeitet
- Die Testergebnisse zu den beiden Zeitpunkten werden als „Items“ X_1 und X_2 aufgefasst
- Es wird die Annahme getroffen, dass die Testergebnisse X_1 und X_2 der beiden Zeitpunkte dem parallelen oder essentiell Modell folgen
- Zur Schätzung der Reliabilität eines einzelnen „Items“ bzw. des Tests kann dann wegen der Parallelität die Korrelation eines “Items“ bzw. Tests mit dem anderen “Item“ bzw. Test herangezogen werden: $REL(X_i) = COR(X_1, X_2)$
- Der große Nachteil dieser Methode besteht wieder darin, dass die dritte Folgerung aus der Parallelität (gleiche Kovarianzen aller Items) nicht überprüft werden kann

Übersicht: Reliabilität des Mittelwerts der Items eines Tests

- Modellspezifische Methoden der Internen Konsistenz (empfohlen):
  - (Essentiell) Paralleles Modell: Spearman-Brown-Formel (vereinfacht), Cronbachs Alpha, McDonalds Omega
  - (Essentiell) τ-äquivalentes Modell: Spearman-Brown-Formel, Cronbachs Alpha, McDonalds Omega
  - τ-kongenerisches Modell: McDonalds Omega

  ➢ Cronbachs Alpha kann auch im τ-kongenerischen Modell angewendet werden, stellt dort jedoch nur eine Mindestschätzung der Reliabilität dar

  ➢ In der Praxis wird unabhängig vom geltenden Modell sehr häufig Cronbachs Alpha verwendet – wie angemessen dieses Vorgehen ist wird kontrovers diskutiert

- Weitere Methoden, die unabhängig vom Modell der Testitems sind:
  - Split-Half-Methode
  - Paralleltestmethode
  - Retest-Methode

Das war‘s schon fast für heute…

- *Ausblick:* In der nächsten Vorlesung beschäftigen wir uns mit dem nächsten Hauptgütekriterium, nämlich der Beurteilung der Validität eines Tests.

- *Aber zuerst:*
  - **Gibt es offene Fragen zur heutigen Vorlesung?**
  - Zur Vertiefung:
    - Aufgaben 5-11 im Übungsblatt 7 zur Reliabilität auf Moodle
    - R Code zu den Beispielen der heutigen VL im Zusatzmaterial auf Moodle
    - Zusatzmaterial zu den math. Beweisen auf Moodle
    - Bühner (2021, S. 115-124)

Die Rekrutierung für den Onlinefragebogen im UK Fragebogenentwicklung läuft noch bis Mittwoch!