

Cronbachs Alpha sinnvoll einsetzen

Dominik J. Leiner, Februar 2016, www.dominik-leiner.de/alpha.pdf

Latente Konstrukte werden im standardisierten Fragebogen meist mittels Fragebatterien gemessen, etwa im Stil von Likert-Skalen. Cronbachs Alpha soll in der Auswertung die Güte der Skala belegen und/oder helfen, „schlechte“ Items zu entfernen. Der folgende Text zeigt Fallstricke beim Einsatz des Kennwerts α .

Cronbachs Alpha und die Reliabilität

Hinter Cronbachs Alpha steht die bestechende Idee vom Split-Half-Test auf Reliabilität. Kann man keine tatsächliche Messwiederholung (Test-Retest) durchführen, bedient man sich einer virtuellen Messwiederholung. Dazu werden die Skalenitems in zwei Gruppen geteilt – schon liegen eine „erste“ und „zweite“ Messung des Konstrukts mit der (jeweils halben) Skala vor, die sich vergleichen (korrelieren) lassen. Diese virtuelle Messwiederholung vermeidet en passant auch das Problem von Erinnerungs- und Entwicklungseffekten.

Cronbachs Alpha optimiert die Berechnung noch ein wenig: Er korreliert alle Items miteinander und verrechnet die durchschnittliche Korrelation \bar{r} zwischen den Items mit einem Korrekturfaktor. Denn je mehr Items (N ist die Anzahl der Items) zur Messung verwendet werden, desto eher heben sich die Messfehler der einzelnen Items auf:

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$$

Cronbachs Alpha ist demnach ein Kennwert, (a) wie viele Items die Skala hat und (b) wie hoch diese durchschnittlich miteinander korrelieren.

Gemessen wird also zunächst die Konsistenz der Skala: Ist die Skala eindimensional oder verbergen sich darin unterschiedliche (Teil-)Konstrukte? Nur wenn Items dasselbe Konstrukt messen, ist es sachlich sinnvoll, die Einzelmessungen zu einem Skalenindex zu verrechnen. In einem zweiten Schritt dient der Kennwert als *Schätzung* für die Test-Retest-Reliabilität der Skala.

In der Literatur wird als Grenze für eine „gute“ Skala oft ein $\alpha \geq 0,7$ angegeben. Diese (willkürliche) Grenze ist an die Reliabilitäts-Interpretation angelehnt. Liefert eine Messung eine Test-Retest-Reliabilität von $r = 0,7$, so ist die erklärte Varianz $r^2 = 0,49$ – oder mit anderen Worten: Der Messfehler bzw. das Messrauschen wird auf die Hälfte der beobachteten Varianz ($\leq 51\%$) beschränkt.

Verzerrungen des Alpha-Werts

Cronbachs Alpha ist nur eine *Schätzung* für die Reliabilität einer Skala. Wichtig für die Interpretation: Einige Rahmenbedingungen können zu größeren Abweichungen zwischen α und der Retest Reliabilität führen.

1. Kurzskalen

Cronbachs Alpha wurde für lange Skalen erdacht und ist sehr sensibel für die Anzahl der Items. Für ein Alpha von 0,7 benötigt man bei drei Items eine mittlere Korrelation von 0,44; bei hundert Items nur noch 0,02. Kann der Befragte drei Items zu einem wenig-latenten Konstrukt (z.B. Servicequalität) gut beantworten, wird die Kurzskala eine hohe Test-Retest-Reliabilität haben – aber Cronbachs Alpha wird kaum 0,7 erreichen.

2. Akquieszenz und andere Messartefakte

Messartefakte können die mittlere Korrelation und damit α „verbessern“. Alleine die Neigung, eher links oder eher rechts auf einer Skala anzukreuzen, sorgen für eine Korrelation der Items untereinander. Im Umkehrschluss bedeutet das: Wenn man die Hälfte der Items negativ formuliert („dreht“), fällt Cronbach's Alpha niedriger aus – bei unveränderter Retest-Reliabilität.

Wichtig ist, dass die Verständlichkeit der Items unter der geänderten Formulierung nicht leidet. Vorsicht daher mit profanen Verneinungen.

3. Substitutive Skalenitems

Cronbachs Alpha ist nur anwendbar, wenn alle Items komplementär dasselbe Konstrukt messen (reflexive Messung). Das klingt banal, ist es aber nicht: Man stelle sich 20 Wissensfragen vor, die das politische Wissen eines Teilnehmers messen. Wissen wird hier als substitutive Größe operationalisiert (formative Messung): Kennt man den aktuellen Wirtschaftsminister nicht, ist es ebenso gut, den aktuellen Außenminister richtig zu benennen. Die Teilfragen *dürften* untereinander gar nicht zu stark korrelieren, denn sonst würde der Wissensindex keine Differenzierung mehr liefern. Bei $\alpha = 1$ wären nur noch 0 oder 20 Punkte möglich.

4. Ungenauigkeit durch die Stichprobe

Die mittleren Korrelation und damit auch Cronbachs Alpha sind mit einer Ungenauigkeit aufgrund der Stichprobenziehung belastet.

Größenordnung: Wird bei 8 Items mit 120 Teilnehmern ein α von 0,65 beobachtet, liegt α zu 95% tatsächlich zwischen 0,5 und 0,8.

Empfehlungen für die Entwicklung (provisorischer) Skalen

Ein fehlerhafter Gebrauch von Cronbachs Alpha bei der Skalengültigkeit kann die Validität einer Messung beeinträchtigen. Folgende Empfehlungen sind daher wichtig für die Validität der Messungen und schränken die tatsächlichen Reliabilität *nicht* ein – sie führen aber meist zu einem „schlechteren“ Alpha.

1. Inhaltsvalidität statt maximaler Konsistenz

Inhaltliche Validität ist dann gegeben, wenn die Items einer Skala alle relevanten Aspekte des Konstrukts abbilden. Sollten die Teilaspekte untereinander nur schwach korrelieren, muss man evtl. mit Teilskalen arbeiten, um das Konstrukt adäquat zu erfassen. Praktischer Nebeneffekt: Man vermeidet, dass sich Teilnehmer durch 10 Paraphrasen derselben Aussage veralbert fühlen.

2. Items „drehen“

Wenn die Hälfte der Items einer Skala negativ formuliert ist (sodass der Teilnehmer durch mehr Zustimmung einen geringeren Indexwert erzielt), dann vermeidet dies Messartefakte durch Akquieszenz – eine wesentliche Ursache für Fehler erster Art in Befragungsstudien.

Im Idealfall ist für die „gedrehten“ Formulierungen keine Verneinung erforderlich, denn diese wird schnell überlesen und kann daher neue Messartefakte verursachen.

3. Intensität der Items variieren

Die Original-Literatur zur Skalenentwicklung (etwa zur Likert-Skala) weist nicht nur auf „gedrehte“ Items hin. Auch sollten die Items einer Skala unterschiedlich intensiv/extrem formuliert sein sollen. Dies gewährleistet eine gute Differenzierung und erhöht die Reliabilität, weil moderate Teilnehmer nicht nur um die Mittelkategorie herum ankreuzen.

4. Items zurückhaltend bereinigen

Man kann berechnen, wie Cronbachs Alpha ausfallen würde, wenn man ein Item aus der Skala entfernt. Dies wird sehr gerne verwendet, um ungeeignete Items in einer provisorischen Skala zu identifizieren – etwa Items, die ambivalent verstanden werden oder eher ein anderes Konstrukt spiegeln. Obwohl grundsätzlich sinnvoll, muss man eine vermeintliche Verbesserung der Skalenkonsistenz kritisch hinterfragen: Geht die Verbesserung über das Messrauschen hinaus? Ist das Item wirklich (nachvollziehbar) problematisch? Verliert man mit dem Item einen relevanten Aspekt des Konstrukts und damit inhaltliche Validität?

Fazit

Cronbachs Alpha ist ein wichtiges Hilfsmittel für die Beurteilung der Skalenkonsistenz und eine sinnvolle *Schätzung* für die Test-Retest-Reliabilität einer Skala, wenn keine Messwiederholung möglich ist. Allerdings müssen bei der Interpretation von α einige Rahmenbedingungen beachtet werden (s. Verzerrungen).

Darüber hinaus verleitet die einfache Ermittlung des Kennwertes dazu, Skalen einseitig in Hinblick auf die Konsistenz der Items zu optimieren – oft zu Lasten der Validität (s. Empfehlungen).

Es hat nach wie vor eine Daseinsberechtigung, die Reliabilität von Skalen in (echten) Messwiederholungen zu ermitteln. Denn so intelligent die Überlegungen hinter Cronbachs Alpha sind ... für die tatsächliche Test-Retest-Reliabilität liefert der Kennwert nur eine für Verzerrungen anfällige Schätzung.

Nicht vergessen ...

1. Gedrehte Items umkodieren/definieren

Je nach Software und Kodierung müssen negativ formulierte Items vor der Berechnung von Cronbachs Alpha ggf. umkodiert (SPSS) oder als gedrehte Items ausgewiesen werden (GNU R).

2. Teilskalen getrennt betrachten

Falls in eine Skala mehrere Teilstrukturen abbildet, muss Cronbachs Alpha für jede Teilskala einzeln ermittelt werden. Die Reliabilität kann nur für die Messung einzelner Konstrukte ausgewiesen werden, nicht für den gesamten Fragebogen.

Zum Nach- und Weiterlesen

Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78(1), 98–104. doi:10.1037/0021-9010.78.1.98

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage. S. 797–806.