



# Fragebogenanalyse

von Daniela Keller

Erheben Sie Daten mittels Fragebogen – egal ob online oder auf Papier – so werden Sie in der statistischen Auswertung mit besonderen Themen konfrontiert. Insbesondere dann, wenn Sie Dimensionen und Konstrukte mittels Items abfragen, wie es häufig in Fragebögen der Fall ist.

Immer dann, wenn ein gewünschtes Kriterium nicht direkt abgefragt werden kann – also wenn es sich um eine latente Variable handelt – werden Items eingesetzt. Das sind Fragen oder Aussagen mit standardisierten Antwortmöglichkeiten. Mehrere solcher Items fragen zum Beispiel die Einstellung zu einem bestimmten Thema ab. Im Anschluss an die Datenerhebung werden die Antworten zu diesen Items für die Auswertung zu einer Variablen zusammengefasst, zum Beispiel als Mittelwert wie in Abbildung 1. So eine zusammengefasste Variable wird oft als Skala, Dimension oder Konstrukt bezeichnet. Gehören mehrere Variablen zu einem übergeordneten Konstrukt, so werden diese Subdimensionen oder Subskalen genannt.

Die Verwendung solcher Items und Dimensionen in der Statistik erfordert den Einsatz bestimmter statistischer Methoden. Dabei kann man grob zwei unterschiedliche Situationen unterscheiden:

Wenn das Ziel Ihrer Untersuchung die Validierung des Fragebogens ist, so können Sie zahlreiche Gütekriterien berechnen und Untersuchungen anstellen. Was hier jeweils möglich und notwendig ist, hängt stark vom Studiendesign und dem Ziel der Studie ab. Die Durchführung einer solchen Validierung soll nicht Inhalt dieses Artikels sein.

Hier sehen wir uns an, was Sie tun, wenn Sie den Fragebogen nicht validieren möchten, sondern ihn lediglich benutzen, um die Variablen zu erheben, die Sie für die Untersuchung Ihrer eigentlichen Forschungsfragen benötigen. Sie werden dann dennoch ein paar wenige Gütekriterien berechnen – nämlich die, die aufgrund des Studiendesigns möglich sind. Außerdem sind weitere Schritte notwendig, damit Sie die Variablen später für Ihre eigentliche Auswertung gut nutzen können. Diese Schritte sind Thema dieses Artikels.

**Untersuchung der Faktor-Item-Struktur**

Je nach Situation kann es notwendig sein, dass Sie die inhaltlich vorgegebene Zuordnung der Items zu den einzelnen Dimensionen bzw. Subdimensionen mittels konfirmatorischer Faktorenanalyse bestätigen möchten. Wenn dagegen die Faktor-Item-Struktur noch nicht feststeht, wenn Sie vielleicht noch nicht einmal wissen, ob es überhaupt Subdimensionen gibt und wenn ja, wie viele, dann führen Sie eine explorative Faktorenanalyse durch.

Mittels Faktorenanalysen wird außerdem ein Teil der Konstruktvalidität, nämlich die faktorielle Validität, untersucht. Es gibt diverse andere Validitätskriterien, zum Beispiel Inhalts- und Kriteriumsvalidität, konvergente und diskriminante Validität. Abseits von speziell für diesen Zweck konzipierten Validierungsstudien sind diese Untersuchungen jedoch meist nicht möglich.

**Gütekriterien der Items und Konstrukte**

Die Items können mehr oder weniger gut zur Messung des Konstrukts geeignet sein. Um herauszufinden, wie die Güte der einzelnen Items ist, berechnen Sie die Itemschwierigkeit und die Itemtrennschärfe für jedes einzelne Item sowie die Reliabilität für das Konstrukt insgesamt.

Die Itemschwierigkeit misst die Zustimmungsrate. Ziel ist es hier, insgesamt in der Stichprobe keine zu extremen Antworten zu haben. In so einem Fall wäre das Item zu extrem formuliert und müsste für eine Folgestudie umformuliert werden. Werte zwischen 0,2 und 0,8 gelten als gute Werte (Bortz/Döring 2006). Im Zuge der Berechnung der Itemschwierigkeit sollten Sie direkt auch Mittelwert und Standardabweichung der einzelnen Items berechnen und berichten.

Die Itemtrennschärfe gibt an, wie gut ein einzelnes Item das Konstrukt insgesamt repräsentiert. Werte größer 0,3 gelten als mittelmäßig, ab 0,5 als hoch (Bortz/Döring 2006). Liegt bei einem Item die Trennschärfe im niedrigen Bereich, so ist dieses Item als schlechter Indikator für das Konstrukt zu sehen.

Als ein Maß der Reliabilität wird die interne Konsistenz für mehrstufige Items in der Form wie in Abbildung 1 meist mittels Cronbachs Alpha für das Konstrukt insgesamt berechnet. Sie gibt an, wie genau das Konstrukt gemessen werden kann. Werte größer 0,9 gelten als hoch, ab 0,8 als mittelmäßig (Bortz/Döring 2006). Bei frühem Forschungsstand oder speziellen Themenbereichen werden allerdings auch niedrigere Werte als ausreichend angesehen (Field 2013).

© Daniela Keller

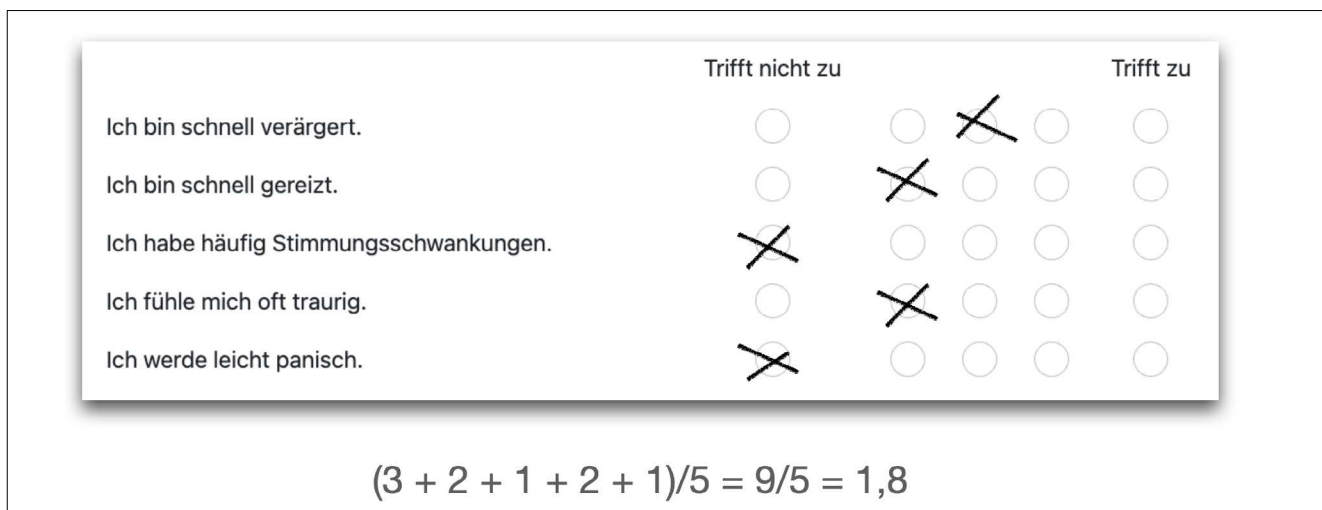


Abbildung 1: Antworten zu fünf Items der Dimension Neurotizismus (in Zeilen). Die fünf Antwortmöglichkeiten (in Spalten) werden mit Zahlen von 1 (trifft nicht zu) bis 5 (trifft zu) kodiert. Die Antworten der Person ergeben als Mittelwert berechnet einen Wert von 1,8 auf der Neurotizismus-Skala.

Für Kurzskalen, die zum Beispiel nur aus zwei Items bestehen, wird die Verwendung der Formel nach Spearman-Brown empfohlen (Eisinga et al. 2012), da Cronbachs Alpha die interne Konsistenz hier unterschätzt.

Sollte Ihr Studiendesign es hergeben, sind möglicherweise auch Berechnungen weiterer Reliabilitätskennwerte, zum Beispiel die Inter-Rater-Reliabilität oder die Retest-Reliabilität möglich. Meist werden diese aber nur in speziellen Validierungsstudien berechnet.

### Skalenbildung

Wenn Sie sich über die richtige Faktor-Item-Struktur klar sind und die Güte der Items und Konstrukte bestätigt haben, bilden Sie nun die Variablen. Das heißt, Sie berechnen sich neue Variablen aus den Items nach einer bestimmten Verrechnungsvorschrift. Meist wird dazu der Mittelwert gebildet. Aber auch ein Summenscore oder ein gewichtetes Mittel sind denkbar. Wenn Sie mit bereits validierten Skalen arbeiten, finden Sie dazu Hinweise im Handbuch. Wenn die Skala neu von Ihnen erstellt wurde, dann dürfen Sie selbst entscheiden. Wichtig ist, dass Sie ihr Vorgehen transparent dokumentieren. Dazu gehört auch die Entscheidung über den Umgang mit fehlenden Werten. Also ob und in welchem Ausmaß Sie fehlende Werte bei einzelnen Items erlauben und damit dann aus den übrigen vorhandenen Angaben die neue Variable bilden dürfen.

Nach der Skalenbildung haben Sie nun in Ihrem Datensatz die neu berechneten Variablen als neue Spalten und können diese wie gewohnt – also genau so, als wären das direkt gemessene Variablen – für die Untersuchung Ihrer Forschungsfrage verwenden. Das heißt, Sie untersuchen die Verteilung, rechnen deskriptive Statistiken, erstellen Grafiken und führen die passenden schließenden Statistiken für Ihre Hypothesen durch.

### Referenzen

Bortz, Jürgen/Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Wiesbaden: Springer.

Eisinga, Rob/Grotenhuis, Manfred te/Pelzer, Ben (2012): The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637–642. <https://doi.org/10.1007/s00038-012-0416-3>

Field, Andy (2013): *Discovering statistics using IBM SPSS statistics*. Los Angeles: SAGE.



© Uschi Matke

### Die Autorin

Daniela Keller ist leidenschaftliche Statistik-Expertin und berät Studierende und Wissenschaftler\*innen zu allen Themen der statistischen Datenanalyse. Während ihres Studiums der Diplom-Mathematik gründete sie mit Kommilitonen eine studentische statistische Beratung und arbeitete anschließend selbstständig in diesem Feld. Neben Einzelberatungen und Workshops unterstützt sie ihre Kund\*innen seit 2019 mit der Statistik-Akademie, ihrem Online-Mitgliederbereich für alle, die Statistik verstehen und selbstständig anwenden wollen. Ihr Blog ([www.statistik-und-beratung.de/blog](http://www.statistik-und-beratung.de/blog)) und ihr YouTube-Kanal sind Fundgruben für leicht verständlich aufbereitetes Statistikwissen für die Praxis.