

# Iterative Ermittlung der Stichprobengröße bei der Faktorenanalyse

Andreas Hinderks<sup>1</sup>, Martin Schrepp<sup>2</sup>, Jörg Thomaschewski<sup>3</sup>

RMT Soft GmbH & Co. KG<sup>1</sup>

SAP SE<sup>2</sup>

Hochschule Emden/Leer<sup>3</sup>

## Zusammenfassung

Fast alle aktuell verwendeten Fragebögen zur Messung von Usability bzw. User Experience basieren auf faktorenanalytischen Verfahren. Bei der Durchführung der Faktorenanalyse ist unter anderem die Größe der Stichproben von Bedeutung. Zur besseren Interpretation der gängigen Empfehlungen über die Stichprobengröße wurde ein Algorithmus entwickelt, der eine iterative Faktorenanalyse durchführt. Die so gewonnenen Ergebnisse der einzelnen Faktorenanalysen wurden untereinander verglichen. Die Ergebnisse zeigen, dass die existierenden Empfehlungen bzgl. der Stichprobengröße praktisch nur sehr begrenzt weiterhelfen.

## 1 Einleitung

Das Ziel einer explorativen Faktorenanalyse ist die Ableitung eines Beziehungsmodells aus einer Menge beobachteter Variablen. Als Ergebnis wird ein Faktorenmodell erzeugt, das mehrere Variablen zu einem Faktor zusammenfasst, bzw. in Beziehung zueinander stellt. Innerhalb der Fragebogenkonstruktion wird die Faktorenanalyse angewendet, um zusammenhängende Items zu Faktoren zusammenzufassen (Bortz und Döring 2006).

Bei der Durchführung einer Faktorenanalyse ist neben der richtigen Auswahl der Items auch die Stichprobengröße von Bedeutung. Zur Beantwortung der Fragestellung, wie groß die Stichprobengröße bei einer durchzuführenden Faktorenanalyse sein sollte, gibt es in der Literatur unterschiedliche Empfehlungen, die sich in zwei Kategorien unterteilen lassen. Empfehlungen der ersten Kategorie basieren auf der absoluten Zahl von Probanden (N), die der zweiten Kategorie auf dem Verhältnis der Anzahl der Items zur Anzahl der Probanden.

Sowohl (Comrey und Lee 2013) als auch (Bühner 2011) propagieren folgende absolute Stichprobengröße (N): 100 = schlecht, 200 = angemessen, 300 = gut, 500 = sehr gut, 1.000

oder höher = exzellent. Bühner (2011) erweitert die untere Grenze auf 60 = gerade ausreichend, sofern die Kommunalität ( $h^2$ ), also das Maß, im welchem die Varianz einer Variable durch alle Faktoren aufgeklärt wird, größer als 0,60 ist. Unterhalb einer Stichprobengröße von 60 ist keine Faktorenanalyse durchführbar (Bühner 2011).

Fabrigar et al. (1999) untersuchte anhand von veröffentlichten Faktorenanalysen die tatsächlich in der Praxis verwendeten Stichprobengrößen. Aus den Ergebnissen lassen sich zwei Schwerpunkte ableiten. Der erste liegt mit einer Stichprobengröße zwischen 100 und 200 (44,2%) und der zweite ist größer als 400 (33,2%).

Die Bandbreite bei der Angabe des Verhältnisses Items zu Probanden ist in der Literatur hoch. Sie reicht von dem Verhältnis 3:1 (Cattell 1978) bis zu 20:1 (Hair 2010). Nunnally et al. (2010) geben die Empfehlung ab, dass das Verhältnis nicht unter 10:1 betragen darf, wohingegen MacCallum et al. (1999) empfehlen 5:1 nicht zu unterschreiten.

Ziel dieses Beitrags ist es zu untersuchen, ab welcher Anzahl von Probanden eine Faktorenanalyse ein stabiles Faktorenmodell ermittelt. Die so ermittelte Stichprobengröße soll mit den vorher beschriebenen Empfehlungen verglichen werden.

Da fast alle aktuell verwendeten Fragebögen zur Messung von Usability bzw. User Experience auf faktorenanalytischen Verfahren basieren, ist diese Frage für die weitere Entwicklung solcher Messinstrumente im Bereich User Experience sehr relevant.

## 2 Methode

Es wurde ein Algorithmus entwickelt, der eine iterative Faktorenanalyse durchführt. Das bedeutet, dass in Schritten von jeweils 10 Probanden eine neue Faktorenanalyse durchgeführt wird. Die Zuordnung der Items zu dem ermittelten Faktor wird anhand der Ladungsgröße  $> 0,45$  vorgenommen (Comrey & Lee 2013). Sofern sich die Zuordnung zwei Mal hintereinander gegenüber den Ergebnissen der vorherigen Faktorenanalyse nicht ändert, wird das Faktorenmodell als stabil angesehen.

Es wurden Datensätze aus 8 verschiedene Evaluation (147 bis 1945 Probanden) mit dem UEQ (Laugwitz et al. 2006) mit dem Algorithmus untersucht. Als Testobjekt wurden verschiedene Produkte untersucht, wie WhatsApp, Hausgeräte aus verschiedenen Produktkategorien, Speicherdienste und eine Monte-Carlo-Simulation).

Der eingesetzte UEQ Fragebogen besteht aus 26 gegensätzlichen Begriffspaaren (z.B. langweilig/spannend, kompliziert/einfach, kreativ/phantasielos), die die User Experience des Testobjektes messen. Die Faktorenanalyse wurde nach der Hauptkomponentenanalyse (Principal components analysis, PCA) mit einer anschließenden Varimax-Rotation durchgeführt.

### 3 Ergebnisse

Tabelle 1 zeigt die Ergebnisse der iterativ berechneten Stichprobengröße der einzelnen Evaluationen, so wie das sich daraus berechnete Verhältnis Item zu Probanden (rechte Spalten). Zudem wurde überprüft, ob das Verhältnis der Items zu der tatsächlichen Anzahl von Probanden mit jeweils 5:1, 7:1 und 10:1 eingehalten wurde.

Evaluation		Faktoren Anzahl	Größe laut Empfehlung erreicht?			Iterativ Berechnet	
Nr.	Probanden		5:1	7:1	10:1	Stabil ab	Tatsächliches Verhältnis
1	1945	2	Ja	Ja	Ja	60	2,3:1
2	500	6	Ja	Ja	Ja	260	10,0:1
3	402	4	Ja	Ja	Ja	350	13,5:1
4	303	4	Ja	Ja	Ja	170	6,5:1
5	293	4	Ja	Ja	Ja	220	8,5:1
6	403	4	Ja	Ja	Ja	240	9,2:1
7	200	4	Ja	Ja	Nein	--	--
8	147	6	Ja	Nein	Nein	--	--

Tabelle 1: Ergebnisse der Faktorenanalyse der Evaluationen

Betrachten wir zunächst die absolute Größe der Stichprobe. Dabei zeigt sich, dass sich bei der Evaluation 1 das Faktorenmodell bereits nach 40 Probanden nicht mehr ändert. Um das festzustellen, musste der Algorithmus jedoch bis 60 Probanden prüfen, also zwei weitere Schritte, um eine Änderung auszuschließen. Die weitere Analyse der restlichen 1.885 Probanden hat keine weitere Änderung des Faktorenmodells ergeben. Bei einer Stichprobengröße von 60 bzw. 40 hätte aber laut Empfehlung (Bühner 2011) keine Faktorenanalyse durchgeführt werden dürfen, was sich in diesem Fall aber als irreführend herausstellt.

Die Evaluationen 2 (ab 260 Probanden), 3 (ab 350 Probanden), 5 (ab 220 Probanden) und 6 (ab 240 Probanden) erzielten jeweils nach 200, bzw. 300 ein stabiles Faktorenmodell. Das entspricht der Empfehlung nach Comrey et al. (2013), was eine *angemessene*, bzw. *gute* Stichprobengröße ist. Bei der Evaluation 4 konnte bereits nach 170 Probanden ein stabiles Faktorenmodell festgestellt werden, was nach Comrey et al. (2013) einer *schlechten* Stichprobengröße entspricht. Interessant ist die Evaluation 7, die eine *angemessene* Stichprobengröße aufweist, aber kein stabiles Faktorenmodell. Für die Evaluation 8 konnte auch kein stabiles Faktorenmodell ermittelt werden.

Betrachten wir nun das Verhältnis Anzahl der Items zu der Anzahl der Probanden. Die Evaluationen 1 bis 6, bei denen ein stabiles Faktorenmodell ermittelt wurde, weisen

unterschiedliche Verhältnisse auf. Bei der Evaluation 1 ist das Verhältnis mit 2,3 am geringsten und mit 13,5 bei der Evaluation 3 am höchsten.

Die Empfehlung 5:1 (130 Probanden bei 26 Items) wurde bei den Evaluationen 1 bis 6 erreicht, hat aber nur bei der Evaluation 1 auch tatsächlich zum stabilen Faktorenmodell geführt. Das Verhältnis 7:1 (182 Probanden bei 26 Items) wurde bei den Evaluationen 1 und 4 nicht eingehalten, hat aber auch hier iterativ zu einem stabilen Faktorenmodell geführt. Ein stabiles Faktorenmodell wurde bei dem Verhältnis 10:1 (250 Probanden bei 26 Items) nur bei den Evaluationen 2 und 3 festgestellt.

## 4 Diskussion

Die Ergebnisse zeigen, dass die existierenden Empfehlungen bzgl. der Stichprobengröße praktisch nur sehr begrenzt weiterhelfen. Wir haben reale Datensätze untersucht und dabei festgestellt, dass in einigen Fällen stabile Faktorenmodelle mit deutlich weniger Daten als in den Empfehlungen beschrieben ermittelt werden konnten. In anderen Fällen reichten auch die maximalen Empfehlungen nicht zur Ermittlung eines stabilen Faktormodells aus. In weiteren Untersuchungen muss geklärt werden, welche Eigenschaften der Daten für diese Variabilität der Ergebnisse verantwortlich sind.

### Literatur

- Bortz, Jürgen; Döring, Nicola (2006): *Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler mit 87 Tabellen*. 4., überarb. Aufl., [Nachdr.]. Heidelberg: Springer-Medizin-Verl. (Springer-Lehrbuch Bachelor, Master).
- Bühner, Markus (2011): *Einführung in die Test- und Fragebogenkonstruktion*. 3., aktualisierte und erw. Aufl. München: Pearson Studium (PS Psychologie). Online verfügbar unter <http://lib.myilibrary.com/detail.asp?id=404890>.
- Cattell, Raymond B. (1978): *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Boston, MA: Springer US. Online verfügbar unter <http://dx.doi.org/10.1007/978-1-4684-2262-7>.
- Comrey, Andrew L.; Lee, Howard B. (2013): *A First Course in Factor Analysis*. 2nd ed. Hoboken: Taylor and Francis. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=1562106>.
- Fabrigar, Leandre R.; Wegener, Duane T.; MacCallum, Robert C.; Strahan, Erin J. (1999): Evaluating the use of exploratory factor analysis in psychological research. In: *Psychological Methods* 4 (3), S. 272–299. DOI: 10.1037//1082-989X.4.3.272.
- Hair, Joseph F. (2010): *Multivariate data analysis*. 7. ed. Upper Saddle River, NJ: Pearson Prentice Hall.
- Laugwitz, Bettina; Schrepp, Martin; Held, Theo (2006): Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: Andreas M. Heinecke und Hansjürgen Paul (Hg.): *Mensch & Computer 2006. Mensch und Computer im StrukturWandel*. München: Oldenbourg, S. 125–134.
- MacCallum, Robert C.; Widaman, Keith F. (1999): Sample Size in Factor Analysis. In: *Psychological Methods* 1999 (Vol. 4. No 1), S. 84–99.
- Nunnally, Jum C.; Bernstein, Ira H. (2010): *Psychometric theory*. 3. ed., Tata McGraw-Hill ed. New Delhi: Tata McGraw Hill Education Private Ltd (McGraw-Hill higher education).

**Autoren****Hinderks, Andreas**

Andreas Hinderks studierte Informatik an der Hochschule Emden/Leer. Danach war er geschäftsführender Gesellschafter der ebis GmbH, die ein branchenspezifisches Warenwirtschaftssystem für den Kücheneinzelhandel entwickelt hat. Seit 2010 ist er als Produkt Manager bei der RMT Soft GmbH & Co. KG beschäftigt. Dort ist er verantwortlich für die benutzerzentrierte und innovative Gestaltung von Geschäftsanwendungen mit Schwerpunkt Prozesssteuerung. Berufsbegleitend studiert er im Masterstudiengang Medieninformatik an der Hochschule Emden/Leer und engagiert sich in der Forschungsgruppe HS Emden/Leer im Bereich der Usability und User Experience.

**Schrepp, Martin**

Dr. Martin Schrepp studierte Mathematik und Psychologie an der Universität Heidelberg. 1990 Abschluss als Diplom-Mathematiker. 1990 – 1993 Promotion in Psychologie. Seit 1994 bei der SAP AG tätig. Bisherige Tätigkeitsfelder waren hier die Konzeption technischer Dokumentation, Software-Entwicklung, User Interface Design und Barrierefreiheit. Hauptinteressen sind die Anwendung kognitionswissenschaftlicher Erkenntnisse auf das Design interaktiver Anwendungen, Barrierefreiheit und die Entwicklung von Methoden zur Evaluation und Datenanalyse.

**Thomaschewski, Jörg**

Dr. Jörg Thomaschewski ist Professor an der Hochschule Emden/Leer mit den Lehr- und Forschungsschwerpunkten Usability und User Experience, Human Computer Interaction, Requirement Engineering, Agile Software Development, Internet-Programming, Markup Languages, E-Learning. Er ist Autor verschiedener Online-Module, u.a. „Mensch-Computer-Kommunikation“, das im Rahmen der Virtuellen Hochschule (VFH) an sechs Hochschul-Standorten eingesetzt wird. Er verfügt über umfangreiche Erfahrungen in Usability-Schulungen, agilen Methoden, IT-Analysen und Beratungen.