Personality and Individual Differences 90 (2016) 326–331

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Factor analysis of geometric figures with four attributes A comparison of PCA, varimax and varimin

Matthias Bellmann

Board Practice Partners, Dickweg 8, D-82057 Icking, Germany

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form 12 November 2015
Accepted 16 November 2015
Available online 28 November 2015

Keywords:
Principal component analysis (PCA)
Exploratory factor analysis (EFA)
Simple structure
Complex structure
Varimax
Varimin
Artificial data

ABSTRACT

It is a common expectation that analytical rotations of factors and components aiming for a simple structure allow dimensions of intercorrelated, manifest variables to be identified with a high degree of certainty. A recently presented counter-model, the rotation to complex structure — supported by a large number of investigations — fundamentally calls this assumption into question. This paper examines the claimed advantage of a rotation to complex structure with the aid of artificially generated data structures whose components were predetermined. For similarity comparisons, it has been possible to show that only a rotation to complex structure provides interpretable and realistic solutions.

1. Introduction

Exploratory factor analysis (EFA) and principal component analysis (PCA) are preferred methods for deconstructing conditions that form the basis of the intercorrelation of variables. PCA is one of the oldest and most widespread multivariate methods (Abdi & Williams, 2010; Fabrigar, Wegener, MacCallum, & Strahan, 1999).[1] While the computation of an analytical solution follows mathematical equations, the assessment of the significance of the factors and components determined depends on the subjective judgement of the researcher.

In order to reduce this subjectivity and provide more unambiguous clarification of how reliably and realistically dimensions and composites are identified by these techniques, a conventional research practice— since Thurstone (1947) at the earliest — is to fall back on artificially generated structures in the form of so-called demonstration analyses: latent factors or components that contribute to the generation of the raw values of a correlation matrix are defined in advance in a data set. The contributions to the correlations are thus controlled to a large extent (Haig, 2014). Examples for this procedure with artificially generated data can be found, for example, in Acito and Anderson (1980), Al-Kandari and Jolliffe (2001), Armstrong (1967), Beauducel (2001), Cattell and Jaspers (1967), Cattell and Sullivan (1962), Davies and Higham (2000), Hardin, Garcia and Golan (2013), Hong (1999), Jolliffe (1972), MacCallum, Widaman, Zhang, and Hong (1999), Overall (1964), Preacher and MacCallum (2003), Revelle (2015), Revelle and Wilt (2013) Schmid and Leiman (1957), Sokal, Rohlf, Zang, and Osness (1980) and Tucker, Koopman, and Linn (1969).

Ertel (2009; 2011a; 2011b; 2011c; 2013a; 2013b) recently proposed replacing the principle of simple structure, which is the aim of most of the rotation methods used as part of EFAs and PCAs, with the principle of complex structure. This is a counter-model whose objective is "to transform the initial factors of an EFA…in such a way that they optimize the common presence of the building blocks found (factors), rather than to aggravate and obscure them" (Ertel, 2011b, p. 43). Ertel demonstrated the usefulness of complex structure modelling (CSM) he developed, which is known as varimin, with the aid of a large number of analyses and re-analyses of real data in a direct comparison with (unrotated) PCA and subsequent varimax rotation.

The work presented here investigates whether the power of a PCA with subsequent varimin rotation claimed by Ertel, compared with that of an initial (unrotated) PCA and the result of a subsequent varimax rotation, also applies to artificially generated and meticulously controlled data sets. To this end, the author has conducted a series of experiments, two of which are explained below.

E-mail address: m.bellmann@boardpracticepartners.com.

[1] For the period 2005 to 2015, ScienceDirect identifies nearly 12,700 papers in journals in which PCA is included in the title, the abstract or the keywords (database query in November 2015).

http://dx.doi.org/10.1016/j.paid.2015.11.032
0191-8869/

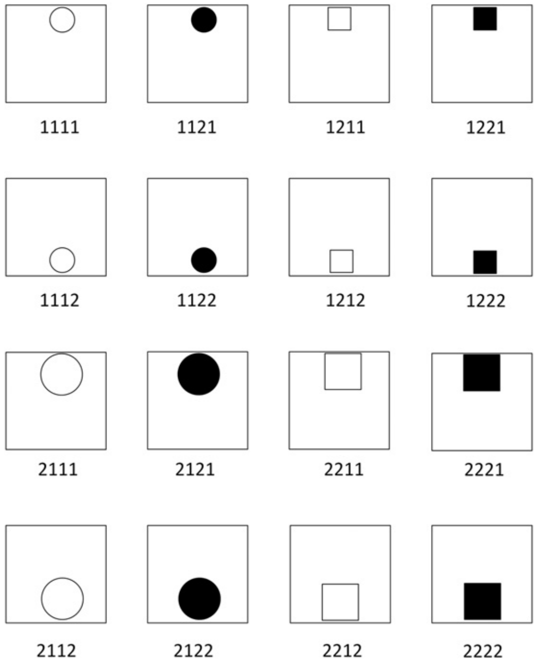

Fig. 1. Sixteen basic geometric forms identified by size, shape, colour and position.

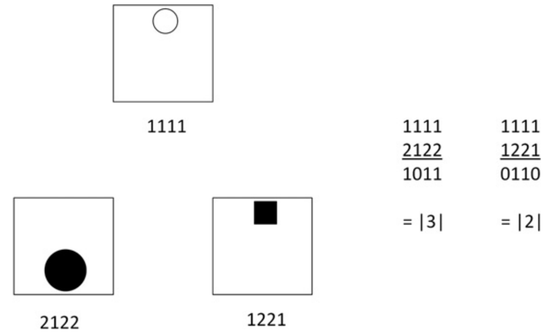

Fig. 2. Example of the triadic similarity determination with the aid of the digits.

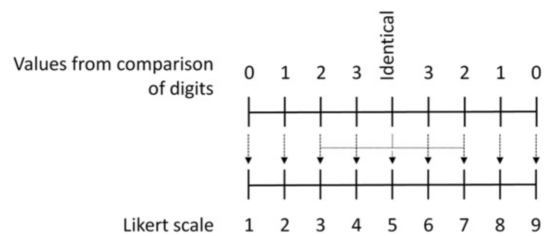

Fig. 3. Transfer of the absolute sum values to a Likert scale.

2. First example

2.1. A triadic comparison

In one of these experimental arrangements, the computer assessed geometric figures differing in size, colour, shape and position (suggested by Künnapas, Mälhammar, & Svenson, 1964; McManus, 1980) for similarity according to the triad method (Sixtl, 1982; Torgerson, 1952; 1967).[2] Two options existed for each attribute, which were encoded as follows:

- 1st digit: small (1) vs. large (2)
- 2nd digit: circle (1) vs. rectangle (2)
- 3rd digit: colour "white" (1) vs. colour "black" (2)
- 4th digit: arranged at top (1) vs. arranged at bottom (2)

Pairing the attributes/options produces sixteen basic forms (see Fig. 1).

120 variants are obtained by combining each basic form with all the others, resulting in a total of 1920 similarity determinations in the triadic comparison.,[3,4] In this experimental arrangement, one can determine by fixed rules which of two basic forms is more similar to a third basic form (the anchoring stimulus). One of the 1920 similarity comparisons is illustrated by way of example (see Fig. 2).

The task is to determine which of the two bottom forms (2122 — left or 1221 — right) is more similar to the top one (1111).

The similarity comparison of 2122 with 1111 results in a (absolute) sum value of "3" after digit-by-digit subtraction; the figures are similar only in terms of their form. Comparing 1221 with 1111 results in the value "2"; the figures are similar in size and position. Hence, the form bottom right has fewer deviations from the top form than the form bottom left. 1221 is therefore more similar to 1111 than 2122, according to the rule introduced.

A nine-point Likert scale was specified for the computer (see Fig. 3); for each of the 1920 comparisons, it had to initially decide whether the (absolute) difference value — illustrated in Fig. 2 only for one comparison — listed on the left or right was lower. The lower value was then transferred to the corresponding point on the Likert scale. If the difference values stated on the left and right were identical, a value from the value range "3 to 7" was selected by a random number generator.

In constellations with identical difference values, which amount to 22.5% of all cases in this experimental setup, this rule gives rise to an error variance when the random selection is not the centre value of the scale (5), but one of the neighbouring values (which of course express a weak preference for one side or the other). Such an arithmetic instruction imitates behaviour that can often be observed in the reality of similarity assessments, where the midpoint of the scale, or a point not too far to the left or right of it, is selected when it comes to a decision under uncertainty.[5]

The 16 profiles (one profile for each basic form), each with $n = 120$ similarity comparisons per basic form, which were generated by the computer on the basis of these specifications, were intercorrelated. The correlation matrix was computed using PCA, and subsequently rotated with both varimax and varimin.[6]

2.2. The findings in detail

The eigenvalues of the PCA suggest a solution with four components: eigenvalues of PC1 to PC5 are 4.14, 3.94, 3.69, 1.48, 0.38, respectively; the explained variance of the first four components is 25.9%, 24.6%, 23.1%, 9.3%, respectively, with a cumulated 82.8%.

Table 1 gives an overview of the three bipolar solutions: PCA, varimin and varimax.

---

[2] Ertel's (2011b; 2013b) coin experiment represents a comparable experimental arrangement, but involved real test persons.

[3] The number of similarity comparisons is given according to the expression $\frac{n!}{k!(n-k)!}n$ for $n = 16$ and $k = 2$.

[4] A mountain of comparisons, which can hardly be expected from real test persons in an experiment.

[5] In this artificial experimental arrangement, error variance is also indispensable because, otherwise, eigenvalues with the contribution of zero would result in matrices that are not positive definite (see Wothke, 1993). The absence of this characteristic can lead to model inconsistencies for the PCA and the subsequent rotation.

[6] The program used was "varimin.exe" (available at www.varimin.com). Its results were crosschecked with the 'psych' package (Revelle, 2015) of the "R" statistics program (R Core Team, 2015), and replicated.