British Journal of Mathematical and Statistical Psychology (2003), 56, 119–125

www.bps.org.uk

A fast method for choosing the numbers of components in Tucker3 analysis

Henk A. L. Kiers and Albert der Kinderen
Department of Psychology, University of Groningen, The Netherlands

Recently, Timmerman and Kiers proposed an effective procedure for choosing the numbers of components in Tucker3 analysis, a kind of component analysis of three-way data. The procedure, however, is rather time-consuming, relying on very many complete Tucker3 analyses. Here, an alternative procedure is proposed, which basically relies on a single, quick analysis of the three-way data set. In a simulation study it was found that the new procedure is comparable in its effect to the original procedure.

1. Introduction

Tucker3 analysis (Tucker, 1966; Kroonenberg & de Leeuw, 1980), also referred to as three-mode principal components analysis, is a method of three-way components analysis of a three-way data set. Such data can, for instance, consist of scores of a number of subjects on a number of variables, measured in a number of different conditions. The aim of Tucker3 analysis is to summarize the three sets of entities constituting the three-way data set in such a way that the main information in the data can be summarized by means of a limited number of components for each set of entities. It is therefore a generalization of two-way principal components analysis (PCA).

In Tucker3 analysis, a (usually preprocessed) three-way data array $\underline{\mathbf{X}}$, of order $I \times J \times K$, is described by the model

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}, \tag{1}$$

$i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$; here a_{ip} and b_{jq} and c_{kr} denote the elements of the components matrices $\mathbf{A}$ $(I \times P)$, $\mathbf{B}$ $(J \times Q)$ and $\mathbf{C}$ $(K \times R)$ respectively, and g_{pqr} denotes the elements of code array $\underline{\mathbf{G}}$ $(P \times Q \times R)$, while e_{ijk} denotes an error term associated with the description of x_{ijk}. The core efficiently describes the main relations

*Requests for reprints should be addressed to Henk A. L. Kiers, Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands (e-mail: h.a.l.kiers@ppsw.rug.nl).

in the data, and the component matrices **A**, **B** and **C** describe how the particular subjects, variables and conditions relate to their associated components.

The model is fitted to a data set by minimizing the sum of squares of the error terms. Specifically, the method minimizes

$$f(\mathbf{A}, \mathbf{B}, \mathbf{C}, \underline{\mathbf{G}}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(x_{ijk} - \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} \right)^2 \qquad (2)$$

over **A**, **B**, **C** and $\underline{\mathbf{G}}$. The 'fit' of the model is calculated as the sum of squares of the data, minus the loss function value $f(\mathbf{A}, \mathbf{B}, \mathbf{C}, \underline{\mathbf{G}})$. It can be verified (see Kroonenberg, 1983, pp. 81–82) that, when the core is the optimal least-squares estimate, the fit is equal to the sum of squares of the approximations to the data, that is, the sum of squares of $\hat{x}_{ijk} = \Sigma_{p=1}^{P} \Sigma_{q=1}^{Q} \Sigma_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr}$. The fit is often divided by the total sum of squares so as to obtain a 'fit proportion', $\Sigma_{i=1}^{I} \Sigma_{j=1}^{J} \Sigma_{k=1}^{K} \hat{x}_{ijk}^2 / \Sigma_{i=1}^{I} \Sigma_{j=1}^{J} \Sigma_{k=1}^{K} x_{ijk}^2$, or 'fit percentage' upon multiplying the fit proportion by 100. For reasons of simplicity, **A**, **B** and **C** are usually constrained to be columnwise orthonormal, which can be done without loss of generality, as it does not affect the optimal fit to be attained. Then the fit is given by the sum of squared core elements. A least-squares algorithm for minimizing (2), and hence maximizing the fit, has been proposed by Kroonenberg and de Leeuw (1980).

An important step in Tucker3 analysis is to choose the numbers of components to use. In contrast to PCA, solutions with different numbers of components cannot be derived from each other. For instance, whereas in PCA the two-components solution for $P = 3$ A-mode components, $Q = 2$ B-mode components and $R = 3$ C-mode components cannot be obtained from one with for instance, $P = 3$, $Q = 3$ and $R = 3$. Therefore, in the case of Tucker3 analysis, for each combination of numbers of combinations, a new analysis has to be carried out. So a strategy could be to carry out all conceivable analyses, compare all results, and then choose the most useful solution, where useful can be defined, for instance, in terms of parsimony, interpretability or stability. However, such an approach of comparing full solutions is practically unfeasible.

To help the user in choosing a sensible set of numbers of components, Timmerman and Kiers (2000) proposed an automatic procedure for selecting the numbers of components to use in a Tucker3 analysis. The procedure is based on the comparison of fit values related to the total number of components used in the analysis, analogous to the principle underlying the scree test (Cattell, 1966). By means of a simulation study, they demonstrated that their 'DIFFIT' procedure worked well in many cases and performed better in recovering the true numbers of underlying components than other selection procedures used until then.

A disadvantage of the Timmerman and Kiers (2000) procedure is that it is rather time-consuming. It computes the fit for Tucker3 models with all sensible combinations of numbers of components, up to a chosen maximum value. Even if for each mode at most six components are taken, this may, even for fairly small data sizes, soon require half an hour or more computation time on computers available nowadays. This is because each analysis consists of an iterative fitting procedure, which, moreover, is to be repeated from different initial configurations a number of times to avoid ending up in local optima. In the present paper, it is proposed to replace these fit computations by a procedure for computing approximate fit values, which has a closed-form solution, and, moreover, gives solutions for all combinations of numbers of components in one go, making it considerably more efficient. This procedure is the one for fitting the Tucker3

model as originally proposed by Tucker (1966). Even though this will lead to suboptimal solutions, we know from experience that the fit is often not very much poorer than that resulting from the least-squares fitting procedure, and moreover, we expect that, when the number of components is increased, the resulting increases in fit values computed with Tucker's fitting procedure will give a good indication of the increases in fit values computed with the least-squares fitting procedure.

In Section 2 Tucker's method for fitting the Tucker3 model will be described. Then, in Section 3, Timmerman and Kiers's (2000) DIFFIT procedure will be summarized and it will be explained how their procedure can be modified so as to use approximate fit values rather than optimal fit values. Finally, in Section 4, a simulation study will be carried out to compare the performance of the DIFFIT procedure applied to optimal fit values with that of the DIFFIT procedure applied to approximate fit values.

2. Tucker's approximate fitting procedure

When Tucker (1966) proposed his three-mode factor analysis model, he also proposed some procedures to estimate the component matrices and the core. These procedures did not give an optimal least-squares fit, but can be considered as approximate least-squares fitting procedures. The simplest of these is based on PCAs of matrices of frontal, lateral or horizontal planes of the array $\underline{\mathbf{X}}$, where the frontal planes are the submatrices associated with the K conditions, the lateral planes are the submatrices associated with the J variables, and the horizontal planes are the submatrices associated with the I subjects. Tucker's procedure is as follows:

Step 1. Define $\mathbf{X}_a$ as the $I \times JK$ matrix with frontal planes of $\underline{\mathbf{X}}$ next to each other. Compute the eigendecomposition $\mathbf{X}_a\mathbf{X}_a' = \mathbf{K}_a\Lambda_a\mathbf{K}_a'$.
Take $\mathbf{A} = \mathbf{K}_{a(1:P)}$ (where $_{(1:P)}$ means taking the columns $1, \ldots, P$ from $\mathbf{K}_a$).

Step 2. Define $\mathbf{X}_b$ as the $J \times KI$ matrix with horizontal planes of $\underline{\mathbf{X}}$ next to each other. Compute the eigendecomposition $\mathbf{X}_b\mathbf{X}_b' = \mathbf{K}_b\Lambda_b\mathbf{K}_b'$.
Take $\mathbf{B} = \mathbf{K}_{b(1:Q)}$ (where $_{(1:Q)}$ means taking the columns $1, \ldots, Q$ from $\mathbf{K}_b$).

Step 3. Define $\mathbf{X}_c$ as the $K \times IJ$ matrix with lateral planes of $\underline{\mathbf{X}}$ next to each other. Compute the eigendecomposition $\mathbf{X}_c\mathbf{X}_c' = \mathbf{K}_c\Lambda_c\mathbf{K}_c'$.
Take $\mathbf{C} = \mathbf{K}_{c(1:R)}$ (where $_{(1:R)}$ means taking the columns $1, \ldots, R$ from $\mathbf{K}_c$).

Step 4. Compute the core as $\mathbf{G}_a = \mathbf{A}'\mathbf{X}_a(\mathbf{C} \otimes \mathbf{B})$, where $\mathbf{G}_a$ is the $P \times QR$ matrix with frontal planes of $\underline{\mathbf{G}}$ next to each other and $\otimes$ denotes the Kronecker product.

Step 5. Calculate the fit proportion as the sum of squared elements of $\underline{\mathbf{G}}$ divided by the total sum of squared elements of $\underline{\mathbf{X}}$.

Step 5 turns out to be very easy to calculate, because the computation of the core (in Step 4) 'happens to' give the least-squares estimate of $\underline{\mathbf{G}}$ given $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Because $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are columnwise orthonormal, the fit is equal to the sum of squares of the core elements. Thus we have a procedure for computing the fit of the estimates computed using Tucker's simple method. This fit, which is not the optimal least-squares fit, will be called the 'approximate fit'. It is expected that using optimal or approximate fit values would not make a lot of difference to the *outcome* of the DIFFIT procedure. On the other hand, the approximate fit value can be computed much more efficiently. This is because the approximate fit values can be computed without using an iterative algorithm, and for different numbers of components this can be done on the basis of a single big core array, which together implies an enormous speed improvement.

To compute fit values for all possible numbers of components, the following procedure can be used. First compute the 'full' core array $\underline{\mathbf{H}}$ associated with the maximal number of components for all three modes, that is, with $\mathbf{A} = \mathbf{K}_a$, $\mathbf{B} = \mathbf{K}_b$ and $\mathbf{C} = \mathbf{K}_c$ so that $\mathbf{H}_a = \mathbf{K}_a'\mathbf{X}_a(\mathbf{K}_c \otimes \mathbf{K}_b)$. Now from this array $\underline{\mathbf{H}}$ we can directly obtain any core array $\underline{\mathbf{G}}$ related to fewer than the maximal numbers of components. This is because the array $\underline{\mathbf{H}}$ has elements $h_{pqr} = \alpha_p'\mathbf{X}_a(\gamma_r \otimes \beta_q)$, where α_p, β_q and γ_r denote the pth column of $\mathbf{K}_a$, the qth column of $\mathbf{K}_b$, and the rth column of $\mathbf{K}_c$ respectively, for $p = 1, \ldots, I$, $q = 1, \ldots, J$ and $r = 1, \ldots, K$. On the other hand, for arbitrary numbers of components (P, Q, R), the core estimate $\underline{\mathbf{G}}$ has elements $g_{pqr} = \mathbf{a}_p'\mathbf{X}_a(\mathbf{c}_r \otimes \mathbf{b}_q)$, where $\mathbf{a}_p$, $\mathbf{b}_q$ and $\mathbf{c}_r$ denote the pth column of $\mathbf{A}$, the qth column of $\mathbf{B}$ and the rth column of $\mathbf{C}$ respectively, for $p = 1, \ldots, P$, $q = 1, \ldots, Q$ and $r = 1, \ldots, R$. Now because $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are taken as $\mathbf{K}_{a(1:P)}$, $\mathbf{K}_{b(1:Q)}$ and $\mathbf{K}_{c(1:R)}$ respectively, it follows that $\mathbf{a}_p = \alpha_p$, $\mathbf{b}_q = \beta_q$ and $\mathbf{c}_r = \gamma_r$, and hence $g_{pqr} = h_{pqr}$, for $p = 1, \ldots, P$, $q = 1, \ldots Q$ and $r = 1, \ldots, R$. Thus, computing the *approximate* fit values of solutions with all possible numbers of components can be done in one go, by computing the full core array $\underline{\mathbf{H}}$ and taking sums of squared elements of parts of this array, the size of which depends on the numbers of components to be taken. Clearly, this is much more efficient than computing new iterative solutions for every different choice of numbers of components.

3. The DIFFIT procedure

Timmerman and Kiers (2000) proposed to choose the numbers of components to use in a Tucker3 analysis by the following procedure, which they called 'DIFFIT'.

Step 1. Collect all fit values for Tucker3 solutions for a large range of combinations of components. For instance, let P, Q and R all assume all integers from 1 to 5, and compute fit values for all 'sensible' combinations. Inserting 'sensible' is meant to exclude cases where $P > QR$, $Q > PR$ and $R > PQ$, because such cases give the same fit as those with $P = QR$, $Q = PR$ and $R = PQ$.

Step 2. Compare solutions with the same total number of components ($s = P + Q + R$) and retain only those solutions that, for a given s, give the best fit.

Step 3. Compute dif_s as the difference between the best solution with s components and that with $s - 1$ components. Determine, from these solutions, the subset of solutions for which $dif_s > dif_{s+j}$, for all $j > 0$. These solutions are indicated $m = 1, \ldots, M$, and the associated numbers of components s are given by $t(m)$, so that their associated *dif* values are given by $dif_{t(m)}$.

Step 4. Compute $b_{t(m)} = dif_{t(m)}/dif_{t(m+1)}$, which gives the ratio of the fit increase resulting from the $t(m)$th component to that of the next interesting component, the $t(m + 1)$th.

Step 5. Select only those numbers of components for which $dif_{t(m)} > \|\underline{\mathbf{X}}\|^2/(s_{\max} - 3)$ to eliminate solutions which led to fit increases that are too small; here $s_{\max}$ is the maximum total number of 'sensible' components. From the remaining solutions, choose that one that has the highest $b_{t(m)}$ value, and denote the associated total number of components as s_c.

Step 6. Choose the numbers of components associated with the best fit among all models using a total of $P + Q + R = s_c$ components.

The rationale for this procedure is described in detail in Timmerman and Kiers (2000). Here, it suffices to mention that the procedure is analogous to Cattell's (1966) scree test.

In the present paper, it is proposed to apply the very same DIFFIT procedure to the approximate fit values calculated by the procedure described in Section 2. Thus, the optimal fit values are replaced by the approximate fit values, and from then on the same procedure is followed.

To illustrate the results of the DIFFIT procedure when applied to the approximate fit values rather than the optimal fit values, we analysed the same data as Timmerman and Kiers (2000) did for illustrating the DIFFIT method. The data set, downloadable from http://three-mode.leidenuniv.nl, consists of scores of six children on five reading tests, administered in 37 consecutive weeks. We followed the same preprocessing procedure as Timmerman and Kiers, and computed all approximate fit values. Next, we applied the DIFFIT procedure to these approximate fit values. In Table 1, the outcomes are reported for the values $s = 3, \ldots, 14$, and for 41, just as was done by Timmerman and Kiers; note that $s = 4$ is not included, because it does not correspond to a set of 'sensible' numbers of components. The outcomes of the Timmerman and Kiers procedure are given in Table 1 as well. It can be seen that the fit values from the two procedures do not differ very much, with the exception of that for $P = Q = R = 1$. More importantly, the dif_s values and the $b_{t(m)}$ values do not differ very much either, and both DIFFIT procedures indicate $P = 2$, $Q = 2$, $R = 1$.

Table 1. DIFFIT results for analyses of learning to read data

	DIFFIT on optimal fit values						DIFFIT on approximate fit values					
$s = P + Q + R$	P	Q	R	fit (%)	dif_s	$b_{t(m)}$	P	Q	R	fit (%)	dif_s	$b_{t(m)}$
3	1	1	1	41.91	41.91	1.49	1	1	1	32.21	32.21	0.87
5	2	2	1	70.12	28.21	4.16	2	2	1	69.16	36.95	5.14
6	2	2	2	76.91	6.78	1.76	2	2	2	76.36	7.19	1.93
7	2	3	2	80.77	3.86	1.27	2	3	2	80.08	3.72	1.26
8	3	3	2	83.82	3.05	1.33	3	3	2	83.03	2.95	1.17
9	3	3	3	86.11	2.29	1.93	3	3	3	85.54	2.51	1.98
10	4	3	3	87.20	1.09	–	4	3	3	86.81	1.27	1.03
11	4	3	4	88.39	1.19	1.06	4	4	3	87.65	0.84	–
12	4	4	4	89.27	0.88	–	5	3	4	88.67	1.01	–
13	5	4	4	90.22	0.95	–	5	4	4	89.90	1.23	1.19
14	5	4	5	91.34	1.12	1.19	5	4	5	90.83	0.93	–
41	6	5	30	100	0	∞	6	5	30	100	0	∞

Note: results set in small font size pertain to dif_s values smaller than the threshold value, in this case 100/38 = 2.63.

4. Simulation study for comparison of DIFFIT based on optimal fit with DIFFIT based on approximate fit

To compare the performance of the original DIFFIT procedure based on optimal fit values to that using approximate fit values, a simulation study was carried out in which 360 data sets were constructed in exactly the same way as in the simulation study by Timmerman and Kiers (2000). The goal of the simulation study was simply to see whether DIFFIT on approximate fit values performed (almost) as well as DIFFIT on

optimal fit values. The performance of the methods was assessed by checking whether or not the method suggested choosing the numbers of components that were actually underlying the constructed data.

We applied both DIFFIT procedures to all 360 data arrays. For DIFFIT on optimal fit, the optimal fit was calculated for all sensible (P, Q, R) combinations up to $P = Q = R = 5$, just as was done in the Timmerman and Kiers (2000) study. For DIFFIT on approximate fit, to keep the comparison fair, the fit was also calculated only for all sensible (P, Q, R) combinations up to $P = Q = R = 5$.

In the 360 analyses, both DIFFIT procedures indicated the correct numbers of components in the great majority (329) of cases. Specifically, DIFFIT on optimal fit did so in 331 cases and, surprisingly, DIFFIT on approximate fit did so even more often: in 336 cases. In two cases DIFFIT on optimal fit indicated the correct numbers of components while DIFFIT on approximate fit did not, in seven cases DIFFIT on approximate fit indicated the correct numbers of dimensions while DIFFIT on optimal fit did not, and in 22 cases both indicated the wrong numbers of components.

These results clearly indicate that replacing the time-consuming DIFFIT on optimal fit procedure by the very quick DIFFIT on approximate fit procedure will not have detrimental effects on the performance of the method. Indeed, in the simulation study, DIFFIT on approximate fit even performed somewhat better than DIFFIT on optimal fit. However, this difference was not found to be significant even at significance level .10 (as assessed by the same repeated multivariate measures analysis of variance procedure as Timmerman and Kiers, 2000, used).

5. Discussion

The present paper offers a very quick alternative to the rather time-consuming DIFFIT on optimal fit procedure. This new procedure, DIFFIT on approximate fit, not only is much quicker but also seems to perform similarly to DIFFIT on optimal fit. In fact, in the simulation study, the new method even performed a little better than DIFFIT on optimal fit, but this difference was small and not statistically significant.

Timmerman and Kiers (2000) advised against using the DIFFIT procedure too rigidly, but rather as a tool for discerning which combination(s) of numbers of components would seem most sensible to select, on the basis of comparisons of fit values of solutions with different total numbers of components. After that, the selected solutions should be studied in more detail to decide which one(s) to maintain. The present results suggest that, in the former step, based on comparison of fit values, it could be sufficient to use approximate rather than optimal fit values.

Acknowledgements

We thank Marieke E. Timmerman for providing us with the software used in the simulation study.

References

Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 174–243). Chicago: Rand McNally.

Kroonenberg, P. M. (1983). *Three mode principal component analysis: Theory and applications*. Leiden: DSWO Press.

Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, **45**, 69–97.

Timmerman, M. E., & Kiers, H. A. L. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology, 53*, 1–16.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.

Received 17 October 2001; revised version received 10 April 2002