

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/62095>

Please be advised that this information was generated on 2019-09-19 and may be subject to change.

ON THE ROTATION OF NON-LINEAR PRINCIPAL COMPONENTS ANALYSIS (PRINCALS) SOLUTIONS: DESCRIPTION OF A PROCEDURE

RUBEN KONIG

In diesem Beitrag wird anhand eines Beispiels gezeigt, wie man in SPSS eine Lösung einer nicht-linearen Hauptkomponentenanalyse (PRINCALS) rotieren kann.

This paper describes an example of how a solution of a non-linear principal components analysis (PRINCALS) can be rotated, using SPSS.

1. Being forced to go beyond ordinary principal components analysis

Ordinary principal components analysis is a technique that can be used to explore the relationships within a group of numeric variables. As such it is a very potent technique, but it has two major drawbacks. a) It can only be used with numerical variables, and b) it can only be used assuming that the relationships between the variables are linear. Therefore, using ordinary principal components analysis to explore relationships that might not be linear, between variables of which the measurement level is uncertain, may be tricky. In such cases, non-linear principal components analysis may help.¹ However, non-linear principal components analysis has some disadvantages, too. One of these disadvantages is the fact that the programme PRINCALS,² that executes non-linear principal components

1 In case the non-linear principal components analysis shows that the variables are all of metric — or near metric— measurement level and that the relationships are linear —or near linear—, one can always choose whether to return to ordinary principal components analysis or not.

2 PRINCALS is incorporated as a procedure in the SPSS package *Categories* (SPSS 1990). However, PRINCALS is not available through the menus of SPSS for windows. It is only available as a syntax command.

analysis by alternating least squares, does not enable rotation to a simple structure. As a consequence, solutions with more than two components are very hard to interpret because they cannot be plotted on paper.

This paper describes how the solution of a non-linear principal components analysis can still be rotated when data were originally intended to serve in an ordinary principal components analysis, but doubts about the measurement level or the linearity of relations have risen. However, the described procedure can only be used when none of the variables in the analysis is treated as a multiple nominal variable.³ The procedure may be used for orthogonal rotation as well as for oblique rotation of the component loadings of the variables. When rotating orthogonally, it is also possible to rotate the object scores (component scores) for data reduction purposes.

Below, you will find a) a very short and incomplete introduction to non-linear principal components analysis, and b) an annotated example of the rotation procedure in SPSS format. The rotation procedure contains variable names that were used in the research of Konig/Renckstorf/Wester (2001). It is evident that, in order to use the procedure yourself, you have to replace these variable names by the names of your own variables.

2. Non-linear principal components analysis

Non-linear principal components analysis differs from ordinary principal components analysis in that it is possible to treat variables not only as numeric, but as ordinal or nominal variables as well.⁴ The categories of all non-numerical variables are assigned a 'category quantification', on a numerical scale. These variables are in fact recoded to give them numerical properties. Of course, the measurement level of the variables restricts the choice of category quantifications. The categories of ordinal variables must be kept in their original order.⁵ The category quantifications make it possible to treat the non-

³ See note 5.

⁴ It is important to note that the choice of measurement level is not only a matter of 'measurement' level, but also of the relationships between variables (Van de Geer 1988; Gifi 1991; De Leeuw 1984). Of course, a nominal variable should not be treated as an ordinal or numeric variable, but it is fairly well possible to treat an ordinal or numeric variable as a nominal variable in the analysis. This, for instance, is needed when the relationship between variables is not monotonic, but curvilinear.

⁵ As to the category quantifications of nominal variables there are two possibilities. It is possible to compute only one set of category quantifications for a nominal variable, or to compute as many sets of category quantifications as there are components in the analysis. In the first case you treat the variable as a 'single nominal' variable, whereas in the second case, you treat it as a 'multiple nominal' variable (Van de Geer 1988; Gifi 1991; De Leeuw 1984).

merical variables as numerical variables and perform a principal components analysis. Where ordinary principal components analysis searches for an optimal mean squared correlation between the original variables and the components, non-linear principal components analysis searches for an optimal mean squared correlation between the variables recoded by the category quantifications, and the components. In the search for an optimal mean correlation between the recoded variables and the components, both the component loadings and the category quantifications are varied until the optimum is found. In ordinary principal components analysis only the component loadings are varied. Readers are explicitly referred to Van de Geer (1988), Gifi (1991), De Leeuw (1984), and SPSS (1990).

3. Annotated example of the rotation procedure in SPSS syntax format

```
COMMENT INPUT VARIABLES FOR THIS ROTATION PROCEDURE ARE THE VARIABLES IN
THE PRINCALS ANALYSIS (SEE BELOW) AND THE RESPONDENT-NUMBER
(RESPNR) .

COMMENT FIVE-DIMENSIONAL NON-LINEAR PRINCIPAL COMPONENT ANALYSIS:
* THE ANALYSIS INVOLVES SOME VARIABLES WITH FIVE, FOUR, AND TWO
  CATEGORIES
* THE ANALYSIS INVOLVES ORDINAL AND SINGLE NOMINAL VARIABLES
* THE UNROTATED OBJECT SCORES (COMPONENT SCORES) ARE SAVED AS
  THE VARIABLES:
  - noroto_1 (object scores on the first component)
  - noroto_2 (object scores on the second component)
  - noroto_3 (object scores on the third component)
  - noroto_4 (object scores on the fourth component)
  - noroto_5 (object scores on the fifth component)
* THE CATEGORY QUANTIFICATIONS, SINGLE CATEGORY COORDINATES,
  AND COMPONENT LOADINGS ARE WRITEN TO THE MATRIX SYSTEM
  FILE 'file_a'.
princals variables=v117 v118 v131 v132 v143 v152 v155 v162 v168 v171
          v184 v228 (5)
          v138 v145 v147 v148 v150 v151 v153 v163 v166 v167
          v170 v172 v175 (4)
          v142 (2)
/analysis=v117 v118 v131 v132 v145 v147 v148 v150 v151 v152
          v153 v162 v163 v166 v167 v170 v171 v172 v175
          v228 (ordi)
          v138 v142 v143 v155 v168 v184 (snom)
/dimension=5
/maxiter=200
/print=eigen quant loadings
/plot=none
/save=noroto
/matrix=out('file_a').
```

```
COMMENT SAVE THE UNROTATED OBJECT SCORES (COMPONENT SCORES) TOGETHER
      WITH THE RESPONDENT NUMBER (respnr) TO THE SYSTEMFILE 'file_b'.
save outfile='file_b'
  /keep=respnr noroto_1 noroto_2 noroto_3 noroto_4 noroto_5.

COMMENT TO ROTATE THE COMPONENT LOADINGS, WE MAKE USE OF THE SPSS COMMAND
      'FACTOR' (THE COMMAND USED TO PERFORM ORDINARY PRINCIPAL
      COMPONENTS ANALYSIS AND FACTOR ANALYSIS), BUT BEFORE WE CAN USE
      THIS COMMAND, WE HAVE TO PREPARE THE UNROTATED FACTOR LOADINGS.

COMMENT FIRST, THE UNROTATED COMPONENT LOADINGS ARE READ FROM THE MATRIX
      SYSTEMFILE 'file_a'.
get file='file_a'.
select if ROWTYPE_="LOADING_".

COMMENT SECOND, THE UNROTATED COMPONENT LOADINGS HAVE TO BE PREPARED FOR
      THE COMMAND 'FACTOR' BECAUSE FACTOR CANNOT READ THE LOADINGS IN
      THE FORMAT THEY ARE IN NOW;
      THEREFORE, ROWS AND COLUMNS HAVE TO BE INTERCHANGED, RESULTING IN
      A MATRIX IN WHICH THE ROWS REPRESENT THE COMPONENTS, AND THE
      COLUMNS REPRESENT THE VARIABLES IN THE ANALYSIS (RECODED BY THE
      CATEGORY QUANTIFICATIONS), WHEREBY:
      * THE NAMES OF THE VARIABLES IN THE ANALYSIS HAVE TO BE READ FROM
        THE VARIABLE 'VARNAME_' IN 'file_a' (NOW ACTIVE FILE)
      * THE COMPONENT LOADINGS OF THE VARIABLES IN THE ANALYSIS ON THE
        RESPECTIVE COMPONENTS HAVE TO BE READ FROM THE VARIABLES 'DIM1'
        TO 'DIM5' IN 'file_a' (NOW ACTIVE FILE).
flip variables=DIM1 DIM2 DIM3 DIM4 DIM5
  /newnames=VARNAME_.

COMMENT THIRD, A NEW VARIABLE 'ROWTYPE_' HAS TO BE CREATED WITH VALUE
      'FACTOR'.
string ROWTYPE_ (a8).
compute ROWTYPE_='FACTOR'.

COMMENT FOURTH, A NEW VARIABLE 'FACTOR_' HAS TO BE CREATED WITH THE
      NUMBER OF THE RESPECTIVE PRINCIPAL COMPONENTS AS VALUE.
compute FACTOR_=$casenum.

COMMENT FIFTH, THE UNROTATED COMPONENT LOADINGS PREPARED IN THIS WAY HAVE
      TO BE SAVED TOGETHER WITH THE NEW VARIABLES 'ROWTYPE_' AND
      'FACTOR_' IN A SEPARATE SYSTEMFILE 'file_c'.
save outfile='file_c'
  /keep=rowtype_ factor_
  v117 v118 v131 v132 v138 v142 v143 v145 v147 v148 v150 v151
  v152 v153 v155 v162 v163 v166 v167 v168 v170 v171 v172 v175
  v184 v228.
```

```

COMMENT NOW THE COMPONENT LOADINGS ARE PREPARED TO BE INSERTED IN THE
      PROCEDURE 'FACTOR' AND STORED IN 'file_c'.
COMMENT ORTHOGONAL ROTATION OF THE COMPONENT LOADINGS USING THE PROCEDURE
      FACTOR'.
factor matrix=in(fac='file_c')
      /criteria=iterate(100)
      /format=sort
      /rotation=varimax.

COMMENT IN CASE YOU ONLY WANT TO INTERPRET THE COMPONENT STRUCTURE, YOU
      ARE READY NOW;
      IF YOU WANT TO USE THE COMPONENTS ANALYSIS FOR DATA REDUCTION,
      YOU WILL ALSO HAVE TO ROTATE THE OBJECT SCORES (COMPONENT
      SCORES), WHICH IS THE AIM OF THE COMMAND LINES BELOW.

COMMENT TO ROTATE THE OBJECT SCORES (COMPONENT SCORES) WE HAVE TO USE THE
      'MATRIX' PROCEDURE OF SPSS.

COMMENT TO BE ABLE TO AFTERWARDS CHECK WHETHER OR NOT WE HAVE MADE
      MISTAKES WE WILL ROTATE THE COMPONENT LAODINGS A SECOND TIME WITH
      THE 'MATRIX' PROCEDURE, WHICH MEANS THAT WE HAVE TO PREPARE AND
      SAVE (IN THE SYSTEMFILE 'file_d') THE UNROTATED COMPONENT LOADINGS
      FOR USE WITHIN 'MATRIX'.
get file='file_a'.
select if ROWTYPE_="LOADING_".
save outfile='file_d'
      /keep=VARNAME_ DIM1 DIM2 DIM3 DIM4 DIM5.

COMMENT STARTING THE PROCEDURE 'MATRIX'.
matrix.

COMMENT CREATING THE MATRIX WITH THE UNROTATED OBJECT SCORES (COMPONENT
      SCORES) (MATRIX NAME: 'noroto').
get noroto
      /file='file_b'
      /variables=noroto_1 noroto_2 noroto_3 noroto_4 noroto_5.

COMMENT CREATING A VECTOR WITH THE RESPONDENT NUMBERS CORRESPONDING TO
      THE UNROTATED OBJECT SCORES (VECTOR NAME: 'respnub').
get respnub
      /file='file_b'
      /variables=respnr.

COMMENT CREATING THE MATRIX WITH UNROTATED COMPONENT LOADINGS
      (MATRIX NAME: 'norotv').
get norotv
      /file='file_d'
      /variables=DIM1 DIM2 DIM3 DIM4 DIM5.

```

```
COMMENT CREATING A VECTOR WITH THE VARIABLE NAMES CORRESPONDING TO THE
      UNROTATED COMPONENT LOADINGS (VECTOR NAME: 'varnames').
get varnames
  /file='file_d'
  /variables=VARNAME_

COMMENT CREATING THE TRANSFORMATION MATRIX FOR ROTATION
      (MATRIX NAME: 'transfor')
      THIS MATRIX HAS TO BE COPIED FROM THE OUTPUT OF THE PROCEDURE
      FACTOR (CELLS ARE SEPARATED BY COMMAS AND COLUMNS BY SEMICOLONS)
      (NOTE THAT IF YOU CONFIGURED SPSS TO PRINT DECIMAL COMMAS
      INSTEAD OF DECIMAL POINTS IN ITS OUTPUT, YOU WILL HAVE TO CHANGE
      THESE DECIMAL COMMAS INTO DECIMAL POINTS).
compute transfor=
  {.52094, .50920, .46218, .24348,-.44322;
  -.48895, .55874, .43577,-.41575, .29326;
  -.58812, .06618, .07917, .79650,-.09511;
  -.25018,-.56194, .54308,-.25321,-.51243;
  -.28474, .32920,-.54341,-.26333,-.66777}.

COMMENT TO CHECK WHETHER MISTAKES WERE MADE, THE COMPONENT LAODINGS ARE
      ROTATED AGAIN (NAME OF RESULTING MATRIX: 'rotv').
compute rotv=norotv * transfor.

COMMENT THE OUTPUT OF THIS MATRIX PROCEDURE HAS TO BE PRINTED AND
      COMPARED WITH THE OUTPUT OF THE PROCEDURE 'FACTOR' (ROWS ARE
      BEING LABELED BY THE NAMES OF THE VARIABLES).
print rotv
  /title="multiplication component loadings and transformation matrix"
  /rnames=varnames.

COMMENT ROTATION OF THE OBJECT SCORES (COMPONENT SCORES) OF THE
      RESPONDENTS (NAME OF RESULTING MATRIX: 'roto').
compute roto=noroto * transfor.

COMMENT SAVING THE RESULTING MATRIX WITH OBJECT SCORES (COMBINED WITH
      RESPONDENT NUMBERS) AS ACTIVE FILE (FOR USE OUTSIDE THE PROCEDURE
      'MATRIX').
save {respnumb,roto}
  /outfile=*
  /variables=respnr comp1 to comp5.

COMMENT TERMINATING THE PROCEDURE 'MATRIX'.
end matrix.

COMMENT THE OBJECT SCORES ARE AVAILABLE NOW ON THE ACTIVE FILE AS THE
      VARIABLES 'comp1' TO 'comp5'.
```

Contact

Dr. Ruben Konig
Department of Communication Science
University of Nijmegen
P.O.Box 9104
6500 HE Nijmegen
Netherlands
Tel.: +31 24 3615789
r.konig@mailbox.kun.nl
<http://baserv.uci.kun.nl/~rkonig>

References

- Geer, J. P. van de, 1988: *Analyse van kategorische gegevens [Analysis of categorical data]*. Deventer, Netherlands: Van Loghum Slaterus.
- Gifi, A., 1991: *Nonlinear Multivariate Analysis (Reprint with corrections)*. Chichester, England: John Wiley & Sons.
- Konig, R./Renckstorf, K./Wester, F., 2001: *On the Use of Television News: Routines in Watching the News*. Pp. 147-171 in K. Renckstorf/D. McQuail/N. Jankowski (Eds.), *Television News Research: Recent European Approaches and Findings*. Berlin: Quintessenz Books. (A previous version of this article was published in *Communications* 23: 505-525).
- Leeuw, J. de, 1984: *Canonical Analysis of Categorical Data (New edition)*. Leiden, Netherlands: DSWO Press.
- SPSS, 1990: *SPSS Categories*. Chicago, IL: Author.