ITERATIVE PRINCIPAL FACTOR ANALYSIS – APPLICATION ON ORDINAL DATA

By

Njatha Moses Mwangi

SCHOOL OF MATHEMATICS

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

UNIVERSITY OF NAIROBI

A PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS IN PARTIAL FULFILLMENT FOR A DEGREE OF MASTER OF SCIENCE IN SOCIAL STATISTICS

©JUNE 2012
Declaration

I, the undersigned, declare that this project is my original work and to the best of my knowledge has not been presented for the award of a degree in any other University.

NJATHA MOSES MWANGI

Reg. No 156/64618/2010

Signature Date

Declaration By Supervisor

This project has been submitted for examination with my approval as supervisor

Mrs ANN WANG’OMBE

School of Mathematics,

University of Nairobi,

P.O. Box 30197, Nairobi.

Signature Date
Acknowledgements:

To God Almighty be all the glory.

I remain indebted to my supervisor, Mrs. Ann Wang’ombe for her very helpful guidance through this project. To Mr. Ndiriti and Mr. Mwaniki for their suggestions and comments that were instrumental in enhancing completion of this work. To all the teaching and non-teaching staff for their support during the last 2 years for the Masters program. To my classmates for their encouragement through the stay at University of Nairobi: Shadrack, Long’asto, Bernadette, Gabriel, Charles and Linda. To my family members for their love and encouragement throughout this course. To Mr. Makau, Program Manager at Lea Toto Community Based Organization, and all the staff for their support.

Any errors and omissions, views and interpretations remain mine and should not be attributed to any of the above mentioned persons, School of Mathematics.
Dedicated to:

This work is dedicated to the late Dr. Rosemary Wangeci Nguti, School of Mathematics. I am a better person today because she believed in me and always provided encouragement towards attainment of this qualification.
Abstract:

This study discusses the procedure of variables reduction that improves assessment of attributes or a subject of interest. Details of how measured variables can be used to obtain a set of observed underlying factors are provided. These factors though considered obvious it is shown how they can be quantified. Two ordinal data sets of data are used to discuss the different considerations of the procedure. Starting with results of principal component analysis the study uses iterative principal factor analysis. Assuming orthogonal relationship of possible sets of factors, Varimax rotation is applied to distribute variation among factors. The results show how a better fit of a common factor model can be obtained with the consideration of changes in individual communalities. The model considered to provide the best fit explains 62.5% of the variation and includes measured variables with communalities more than 0.5.
1.0 Introduction

1.1 Background

1.2 Problem statement

1.3 Objectives
   1.3.1 Primary objective
   1.3.2 Secondary objectives

1.4 Justification and Significance of the study

2.0 Literature review

3.0 Methodology
   3.1 The Common Factor Model
List of Tables

4.1 Dataset description................................................................................................................24
4.2 Pearson’s Correlation Coefficients between the sub-domains of Child Status Index......30
4.3 Principal components (proportions of variance): Unreduced.................................30
4.4 Summary of dataset used in discussing factor analysis process...............................33
4.5 Factor loadings for a 12 factors fitted model (un-rotated)......................................39
4.6 Proportion and cumulative variances for the 12 fitted factor....................................39
4.7 Factor loadings for a 12 factors fitted model (rotated).............................................40
4.8 Proportion and cumulative variances of 12-factor loadings (rotated)......................40
4.9 Factor loadings for a 5 factors fitted model (rotated)................................................42
4.10 Proportion and cumulative variances of 5-factor loadings fitted..............................42
4.11 Factor loadings for a 7-factors fitted model (rotated)..............................................43
4.12 Proportion and cumulative variances of 7-factor loadings fitted..........................43
4.1 Factor loadings for a 8-factors fitted model (rotated)................................................44
4.2 Proportion and cumulative variances of 7-factor loadings fitted..............................44
List of Figures

4.1 Scatter matrix between care, education, shelter and social domains of Child Status Index.................................................................25

4.2 Scatter matrix between performance, wellness, abuse and emotional status........26

4.3 Scatter matrix between food, health care and nutrition........................................27

4.4 A screen shot of factor analysis on Rcmdr package on R environment...............29

4.5 Scatter matrix plot between variables x2, x3, x5, x6 and x7.................................34

4.6 Scatter matrix plot between variables x7, x8, x9, x10, x11 and x12.....................35

4.5 Scatter matrix plot between variables x13, x15, x16, x17 and x18.......................36

4.6 Scatter matrix plot between variables x19, x21 and x22.................................37

4.7 Scree plot of the principal components.........................................................38
Chapter 1: Introduction

1.1 Background

In providing care to children health care programs monitor different indicators of their well being. These include social and psychological aspects that are correlated to the health status of a child. Over time different models and indices have been used to measure the status of a child's health, quality of life e.t.c. Body Mass Index (BMI) and Quality of Life Index (QOL)

The challenge of scoring is that there no single function for the domains that would suggest the child status at the time of assessment. The attempts to develop measures that consider most of the child’s health, social and psychosocial variables have been limited. The child Status Index (CSI) tool is one such instrument. It’s not clear as to whether it is appropriate for all the ages and is also silent on special statuses as HIV infection. There has been doubt raised on consistency of the scores obtain in CSI with other measurement of assessment of a child’s wellbeing. The time taken before a re-assessment in practice is long, in some cases six months or more. There is therefore no doubt that important changes in a child could be identified when it is too late. Due to differences in judgments, users of the CSI tool may allocate different scores.

The interest of this study was to discuss how factors that may not be observed could be generated and quantified for better scoring of a child's welfare and or other qualitative outcomes that require measurements.
1.2 Problem Statement:

The measurement of child welfare status using data collection tools that are not standardized and obtained results difficult to analyze limits availability of information for evidenced decision making on priorities and interventions to improve the general well-being of children. A better understanding of the underlying factors that affect observed welfare indicators enhances effective programming of development projects. Resources utilized in an optimal way and services are integrated to deal with the real challenges.

1.3 Research Objectives:

1.3.1 The *primary objective* is to discuss how to obtain a combination of unobserved variables that would best measure attributes of interest related to available observed ordinal data using iterative principal factor analysis method.

1.3.2 The *secondary objectives are:*:-

1. Determine whether the data sets considered for this study are appropriate for factor analysis.

2. Describe how IPF Analysis can be used to reduced a set of ordinal variables to small number of unobservable variables.

3. Determine the best fit of a common factor model to the set of ordinal data while discussing the aspects of the model considered.
1.4 Justification and significance of study:

Models that can be used reliably are effective in helping respond to priorities in the area of health programming. Measurements are more objective. In management of a child’s welfare it is important to understand the connection between domains like nutrition, treatment, psychosocial support e.t.c. Best models, therefore, should capture most of the variables under these domains.
Chapter 2: Literature Review

This chapter focuses on available literature that discusses the aspects of a good measuring tool for quality data. Exploration of aspects of validity, reliability and extensive application data collection tools is conducted. An example of Child Status Index is provided. Application of factors analysis in modeling is also highlighted.

In the recent past development projects have created a big interest in the showing evidence based programming. Most governments and implementing partners of community projects have made initiatives in building the capacity to monitor and evaluate their programs and projects. This in essence focuses on measuring of results to inform future programming. However, this remains a challenge for most. Kothari (2005) notes that measurement is a relatively complex and demanding task, especially when it concerns qualitative or abstract phenomena. Further, there is less confidence in accuracy of the results when the tools are not standardized. This means that usage of such tools varies from one researcher or enumerator to the other.

Errors in measurements greatly affect quality of the data and therefore information processed. Discussing the sources of error in measurement, Kothari lists the possible sources of error in measurement as a) Respondent, b) Situation, c) Measurer and d) Instrument. Specifically, he states the following about error associated with the measurer:

The interviewer can distort responses by rewording or reordering questions. His behavior, style and looks may encourage or discourage certain replies from respondents. Careless mechanical processing may distort the findings. Errors may also creep in because of
incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data-
analysis stage.

Given the need to have quality measurement of results or characteristics of subjects of a study
and or for regular monitoring of variables of interest it is important to maintain soundness of
measurement. He suggests that sound measurement should meet the test of validity, reliability
and practicality as the major considerations in evaluating a measurement tool. He goes further
and defines the three aspects as: Validity as extent to which a test measures what it was actually
wished to measure; Reliability as accuracy and precision of a measurement procedure;
Practicality as concerned with a wide range of factors of economy, convenience, and
interpretability. In case of test of validity the following is outlined:

Validity is the most critical criterion and indicates the degree to which an instrument
measures what it is supposed to measure. Validity can also be thought of as utility. In
other words, validity is the extent to which differences found with a measuring
instrument reflect true differences among those being tested. But the question arises: how
can one determine validity without direct confirming knowledge? The answer may be
that we seek other relevant evidence that confirms the answers we have found with our
measuring tool. What is relevant, evidence often depends upon the nature of the research
problem and the judgment of the researcher.

In this excerpt the interest is underlining the notion that the nature of the judgment of the
researcher (Here any enumerator or data collector is considered) has an implication on the
evidence, mostly so when we consider the differences observed in measurement.

An important question that one should therefore ask in relation to validity of a measurement tool
is whether or not the tool is applicable to all situations. In other words a tool should be versatile
enough for adaptability to the different contexts of geography, subject characteristics like age, vulnerability to a condition that affect a characteristic and or terminal pre-existence conditions like HIV+ status or cancer. Bridget and Cathy (2005) referred to the work by Prof. Kelvyn Jones of School of Geographical sciences, University of Bristol in UK, whose insights are fit in discussing this question. He explores the nature and extent of place effects and thereby challenges the familiar critique of quantification that in pursuing generality it ignores specificity. He sets the philosophy:

\[
\text{Outcomes} = \text{Mechanism} + \text{Context}
\]

Simply stated use of measuring tools in social science is dependent of the context in which it is embedded. Jones points out that there is a recursive relationship between people and places such that geography matters so much that human processes cannot be understood without being informed by a geographical imagination.

These ideas are useful in explaining the interest in this study to discuss the process of factor analysis. Given a set of measured/manifest variables would it be possible to generate unobserved variables that could be quantified. Now, many tools have been developed to measure qualitative aspects of an individual welfare and program outcomes. Here attention is given to CSI.

Child Status index tool (see CSI tool indexed) was developed to measure a child’s wellbeing on the following sub-domains: food security, nutrition and growth, shelter, care, abuse and exploitation, legal protection, wellness, health care services, emotional health, social behavior, performance and education/work.
The need to have a tool that is able to capture a child’s welfare on many aspects for assessment of vulnerability, measurement of outcomes and evaluation of impact led to development of CSI.

As it is the expected process in development of measuring tools practical testing for the validity, reliability and practicability should be assessed based on feedback from the end users while comparing with other existing tools known to be considered conventionally as best practice. Lora Sabin et al set out to validate the CSI tool as an instrument that can meaningfully measure vulnerabilities of Orphans Vulnerable Children (OVC) including those infected and affected by HIV and AIDS. Studies to validate CSI have been conducted by different research teams. The following is one of the results:

No relationships exceeded the standard for high construct validity ($0.7). Only 2 were moderate (0.3–0.7), both for the younger age group: food security (0.4) and wellness (0.36). All other relationships were weak or negative. In most subcategories, a substantial proportion of surveyed children indicated distress that was not evident from CSI scores. In the abuse and exploitation sub-domain, all children were rated as “good” or “fair” by the CSI, but among surveyed children aged 11–17, 20% or more reported being beaten, kicked, locked out of the house, threatened with abandonment, cursed, and made to feel ashamed.

The team of researchers from Center for Global Health and Development; department of International Health, School of Public Health, Boston University, Boston, Massachusetts; and The Centre for Social Research, University of Malawi, Zomba, Malawi, concluded that they were unable to validate the CSI as a tool for assessing the vulnerabilities of orphaned and vulnerable children in rural Malawi population of study. They recommend caution in interpreting CSI scores and revisions to the tool before global scale-up in its use.
positively or negatively) are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors.

Elizabeth (2006) discusses the procedures for conducting factor analysis noting that it is mostly confused with principal component analysis. Here weights and factor loadings are differentiated. Some of the challenges in interpretation of the results are highlighted as naming of factors and concluding that factors represent separate groups of people. The criticisms of factor analysis as suggested are: labels of factors can be arbitrary or lack scientific basis, derived factors often very obvious but a quantification is achieved, poor input variables result in poor outputs, the process is too complicated, many steps are engaged and could affect results and that the correlation matrix is often a poor measure of association of input variables.

Field (2005) notes factor analysis could be problematic in cases where values of pearson's correlation matrix are greater than 0.9. Singularity of the system could arise implying that the inverse would not exist. It is suggested that elimination of one of the two variables that return the high correlation value would deal with this problem. A necessary value of 0.00001 is set for making a decision on admissible determinant of the matrix system. This information is important in identifying the problem of multicollinearity. Using SPSS, Field (2005) notes that there are as many eigenvectors as there are variables and so there will be as many factors as variables. The eigenvalues associated with each factor represent the variance explained by that particular linear component. Factors are extracted and then rotated. Rotation has the effect of optimization of the factor structure with the consequence of equalizing the relative importance of selected factors. Different techniques of rotation tend to produce similar results for large samples and well
defined factors. Rcmdr uses the varimax (orthogonal) method which attempts to minimize the number of variables with high loading on a factor enhancing interpretability of the factors (Luiz and Graeme 2010). Orthogonal rotation is preferred for theoretically independent factors. Further, it is stated that Principal component analysis works on the initial assumption that all variance is common. Upon extraction however, communalities obtained are looked at in terms of the proportion of variance explained by the underlying factors. Discarding of some factors after extraction implies that some information is lost.

Determination of the appropriateness of running a factor analysis can be conducted using KMO (Kaiser Meyer Olkin) statistics as measures of sampling adequacy. These can be calculated for individual and multiple variables using equations.

Results output of a factor loadings matrix can be requested to include values equal to or greater than a set value, say 0.4, with others suppressed. The number of factors to retain is the decision of the researcher. Kaiser’s criterion is noted to be accurate when there are less than 30 variables and communalities after extraction are greater than 0.7 or when the sample size exceeds 250 and the average communality is greater than 0.6. Research into Kaiser’s criterion gives recommendations for much smaller sample, it suggested therefore that huge sample produce more accurate results. A scree plot can also be used (when the sample size is about 300 or more) to decide the number of the factors considering the point of inflexion on the curve. If generated factors represent real-world construct then common themes among highly loading variables are useful in identifying what the construct might be.
The choice of factor model is also preferred because it can cope with many variables without running into scarce degrees of freedom problems that characterize regression-based analysis (Jorg and Sandra 2005). In the next chapter the methodology for fitting such a model is detailed.
Chapter 3: Methodology

3.1 The Common Factor Model:

Consider a data model operating in a population of a large N defined typical as follows

\[\begin{align*}
X_j &= \mu_j + \lambda_{j1}z_1 + \lambda_{j2}z_2 + \ldots + \lambda_{jm}z_m + u_j + \epsilon_j \\
&= \mu_j + \sum_{k=1}^{m} \lambda_{jk}z_k + u_j \quad j = 1, \ldots, p
\end{align*}\]

Where \(X_j\) is the score for a typical research subject on measured variable \(j\)

- \(\mu_j\) represents mean of measured variable \(j\)

- \(z_k\) common factor score for a typical research subject on factor \(k\)

- \(\lambda_{jk}\) factor loading of test \(j\) on factor \(k\)

- \(u_j\) is the factor score for a typical research subject person on a unique factor \(j\)

- \(S_j + e_j\) is the factor score for a typical research subject on specific factor \(j\)

- \(e_j\) is the error term for a typical subject on test \(j\)
Model (1) is expressed in matrix notation where \( x = \mu + \Lambda z + u \)

\[
x = \begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_p
\end{bmatrix} \quad \text{typical score of measured variables}
\]

\[
\mu = \begin{bmatrix}
    \mu_1 \\
    \mu_2 \\
    \vdots \\
    \mu_p
\end{bmatrix} \quad \text{mean of measured variable } j
\]

\[
\Lambda = \begin{bmatrix}
    \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1m} \\
    \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    \lambda_{p1} & \lambda_{p2} & \ldots & \lambda_{pm}
\end{bmatrix} \quad \text{pxm (p > m) factor loadings}
\]

\[
z = \begin{bmatrix}
    z_1 \\
    z_2 \\
    \vdots \\
    z_m
\end{bmatrix} \quad \text{unobserved common factor scores}
\]

\[
u = \begin{bmatrix}
    u_1 \\
    u_2 \\
    \vdots \\
    u_p
\end{bmatrix} \quad \text{unobserved unique factor scores}
The only available data for system (2) above is the measured MVS. This is usually the case in a practical case scenario. Therefore, a transformation is necessary to obtain a solvable system simultaneously.

3.2 Covariance structure of the data model (1):

We assume

1. The common factors \( z \) and unique factors \( u \) are independently distributed. Consequently, common factors are uncorrelated with unique factors,
   \[
   \sum_{zz} = 0 = \sum_{uz}
   \]

2. The \( P \) unique factors forming elements of \( u \) are mutually independent. The implication of this is that \( \sum_{uu} \) is diagonal

3. We suppose that common factors \( z \) and unique factors \( u \) are standardized with mean zero vectors for identification process.

4. Further suppose those common factors \( zs \) are standardized to have unit variances.
**Notation - wise presentation**

Measured variable covariance matrix \( \sum = \sum_{xx} \)

Common factor covariance matrix \( \phi = \sum_{zz} \)

Unique factor covariance matrix \( D_{\psi} = \sum_{uu} \)

Given assumption (4) \( \phi \) is a factor correlation matrix.

Consider the matrix form of the data model (1)

\[
x = \mu + \Lambda z + u
\]

Assumption (3) implies

\[
E[x] = E(\mu) + E[\Lambda z] + E(u)
\]

\[
\mu_x = \mu + \Lambda(0) + (0)
\]

This result suggest that mean of the variables is unrestricted by the model (i.e. is the constant vector \( \mu \))

To obtain \( \sum_{xx} \) again using matrix form (1)

\[
x - \mu = \Lambda z + u \quad \text{(3)}
\]
Equation (3) expresses deviation of individual scores from population means as a linear function of factor scores and factor loadings.

\[(x - \mu)(x - \mu)' = (Az + u)(Az + u)'
\]
\[= (Az + u)(Az)' + u'u'
\]
\[= (Az + u)(z'\Lambda' + u')
\]
\[= \Lambda zz'\Lambda' + \Lambda z u' + uz'\Lambda' + uu'
\]

Note: Matrix properties and operations are not discussed in this study.

Taking the expected values on both sides we have

\[E(x - \mu)(x - \mu)' = \Lambda E(zz')\Lambda E(zu') + E(Uz')\Lambda' + E(uu')
\]
\[\sum_{xx} = \Lambda \sum_{zz} \Lambda + \Lambda \sum_{zu} + \sum_{uu} \Lambda' + \sum_{uu}
\]

Simply written as

\[\sum_{xx} = \Lambda \phi \Lambda' + D_{\psi} \ldots (4)
\]

Note: assuming first that the common factors are uncorrelated we obtain.

\[\sum_{xx} = \Lambda \Lambda' + D_{\psi} \quad Since \phi = I \ldots (5)
\]

From equation (4) above we express the observed variance for each $MV_{x_j}$ as

\[\delta_{jj} = (\Lambda \phi \Lambda')_{jj} + \Psi_{jj}
\]

$\Psi_{jj} -$ Unique variance of $x_j$
The communality of $MV_{x_j}$, denoted by $h_{ij}$ is given by

$$h_{ij} = \frac{[\Lambda \Phi \Lambda^T]}{\delta_{ij}} = 1 - \frac{\psi_{ij}}{\delta_{ij}} \ldots (6)$$

3.3 Correlation structure:

Conveniently we shall work with the correlation matrix. Consider the diagonal matrix (elements are variances of the measured variables)

$$D_{\delta} = \text{Diag} \left| \sum \right|$$

We obtain $P$, population correlation matrix as follows:

$$P = D_{\delta}^{-\frac{1}{2}} \sum D_{\delta}^{-\frac{1}{2}}$$

$P$ has diagonal entries of 1.0 and off-diagonals as correlation coefficients among variables.

Now we have the factor analysis correlation structure given as

$$P = D_{\delta}^{-\frac{1}{2}} (\Lambda \Phi \Lambda^T + D_{\psi}) D_{\delta}^{-\frac{1}{2}}$$

$$= \Lambda \Phi \Lambda^T + D_{\psi}^* \ldots (7)$$

Where;

$$\Lambda^* = D_{\delta}^{-\frac{1}{2}} \Lambda$$

$$D_{\psi}^* = D_{\delta}^1 D_{\psi}$$
Note 1: Unique variances for the correlation structure actually correspond to proportions of variance of each measured variable attributable to unique factors.

Note 2: if two measured variables have high loadings on one or more of the same factors, the sum of products of loadings will tend to be high. If two measured variables do not load highly on any of the same factors, the sum of products of loadings will be low. In this, sense, the common factors account for the correlations among the measured variables.

Note 3: An off-diagonal element of the population correlation matrix P with the corresponding entry in $D^*_y$ as zero implies that the unique variances do not contribute to explaining correlations between MVs.

3.4 Rotational Indeterminacy of the Factor Matrix:

If a single $pxm$, $m \geq 2$, factor matrix $A^*$, can be found so that $P = A^*_1A^*_1 + D_y$, and if $m \geq 2$, then there are infinitely many other $pxm$ factor matrices, $A_2, A_3 ...$ such that

$$ P = A^*_2A^*_2 + D^*_y = A^*_3A^*_3 + D^*_y = ... $$

If $P = A^*_1A^*_1 + D_y = A^*_2A^*_2 + D^*_y$ then $A^*_2 = A^*_1T$ Where $T$ is a $mxm$ orthogonal matrix (e.e. $TT^T=1$).

Therefore, to achieve a unique solution, one way is using eigenvectors and eigenvalues of $P$.

The eigen-structure of any symmetric matrix $S$ is given by
\[ S = UD_i U^1 \ldots \text{(8)} \]
\[ U \text{ - eigenvectors} \]
\[ D_i \text{ - Eigenvalues as the diagonal elements} \]

Now, if we take \((P - D^*_\psi)\) to be a symmetric matrix, a factor matrix \(\Lambda\) may be constructed from the largest eigenvalues, \(l_1, l_2 \ldots l_m\) and corresponding standardized eigenvectors, \(u_1, u_2, \ldots u_m\) of \((P - D^*_\psi)\).

Note: The \((p-m)\) smallest eigen values of \((P - D^*_\psi)\) are all =0

Since \((P - D^*_\psi) = \Lambda^* \Lambda'^1 = UD_i U^1 \text{ from equation (8)}\)

\[ \Lambda^* \Lambda'^1 = UD_i^{1/2} * D^1_i^{1/2} U^1 \]
\[ \Rightarrow \Lambda^* = UD_i^{1/2} \]

This provides the factor loading matrix.

That is, each column \(\Lambda^*\) is equal to an eigenvector multiplied by the square root of the corresponding eigenvalues i.e.

\[ \lambda_j = u_j \sqrt{l_j} \quad j = 1, \ldots, m \]

There exist \(\frac{1}{2} m(m - 1)\) equations known as identification conditions to be considered that satisfy \(P - D^*_\psi = \Lambda^* \Lambda'^*\)

NB: \(\frac{1}{2} m(m - 1)\) is finite as long as \(m\) is finite.
This therefore, tends to deal with the question of rotational inter-determinacy by reducing the set of possible $\Lambda^*$.

3.5 Estimating communalities:

Estimated values of communalities can be obtained to replace the unique variances in $(P - D_\Psi)$ assuming equation (7). Estimation can be conducted by the method of

(a) Squared Multiple Correlations
(b) Partitioning Method

Since $P$ is mostly unknown in practice, we use an estimate of $R$, sample correlation matrix.

Therefore model (1) will rarely fit exactly in the population or sample

$$R = \hat{\Lambda}^* \hat{\Lambda}^{*\dagger} + \hat{D}_\Psi^* \ldots (9)$$

The degree of approximation is reflected in the residual matrix i.e.

$$R - (\Lambda^* \Lambda^{*\dagger} + D_\Psi^*)$$

At this point if we applied the computations above for population correlation matrix $P$ on $R$, and let

$$\hat{\Lambda}^* = \hat{U} \hat{D}^{1/2}$$

Then $\hat{\Lambda}^*$ minimizes the residual sum of squares

$$\text{RSS} = \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left[ (R - \hat{D}_\Psi^*) - \hat{\Lambda}^* \hat{\Lambda}_j^* \right]^2 \ldots (10)$$
Minimization is done conditional to a given set of prior communality estimates.

This is termed as Principal Factor Method using Prior Communality Estimates

Given \( \hat{\Lambda} \) estimates of the communalities are obtained by:

\[
\hat{h}_{jj} = \sum_{l=1}^{n} \hat{\lambda}_{jl}^2, \quad j = 1, \ldots, P
\]

While the scaled unique variances are given by:

\[
\hat{\Psi}_j^* = 1 - \sum_{l=1}^{n} \hat{\lambda}_{jl}^2, \quad j = 1, \ldots, P
\]

These communality (scaled unique variance) estimates conditionally minimize equation (10)

A procedure useful in estimation of RSS would be

1. Obtain initial communality estimates using SMCs.
   
   Insert these into the diagonal of \( R \) to obtain \( \left( R - \hat{D}_\psi^* \right) \)

2. Obtain \( \hat{\Lambda} \) by the principal factors method, from the eigen values and eigenvectors of \( \left( R - \hat{D}_\psi^* \right) \)

3. Obtain new communality estimates by computing sums of squares of the loadings in each row of \( \hat{\Lambda}^* \). These are considered as new estimates and are inserted into the diagonal of \( R \)
   
   to obtain new \( \left( R - \hat{D}_\psi^* \right) \)
4. Go back to (2) and continue to process until convergence has occurred.

3.6 Goodness of fit:

Let $F_{\text{OLS}}(R, \hat{P}) = \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} [R_{ij} - \hat{P}_{ij}]^2$

Recall: $\hat{P} = \hat{\Lambda} \hat{\Lambda}^* + \hat{D}_w^*$ for uncorrelated factor loadings.

The values $\hat{D}_w^*$ and $\hat{\Lambda}$ are known as ordinary least squares (OLS) estimates as they minimize RSS. The value of this function ($F_{\text{OLS}}$) indicates degree of lack of fit of the model.

OLS communality estimates greater than one are not admissible (Heywood Case)
4. Go back to (2) and continue to process until convergence has occurred.

3.6 Goodness of fit:

Let $F_{OLS}(R, \hat{P}) = \frac{1}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} (R_{ij} - \hat{P}_{ij})^2$

Recall: $\hat{P} = \hat{\Lambda}^* \hat{\Lambda}^{*\prime} + \hat{D}_p$ for uncorrelated factor loadings.

The values $\hat{D}_p$ and $\hat{\Lambda}$ are known as ordinary least squares (OLS) estimates as they minimize RSS. The value of this function ($F_{OLS}$) indicates degree of lack of fit of the model.

OLS communality estimates greater than one are not admissible (Heywood Case)
Table 4.1: Dataset description

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Choice criteria</th>
<th>Recoded variable name</th>
</tr>
</thead>
<tbody>
<tr>
<td>shop</td>
<td>Supermarket customer chooses</td>
<td>shop</td>
</tr>
<tr>
<td>Q08a</td>
<td>Amount of non-grocery products</td>
<td>x1</td>
</tr>
<tr>
<td>Q08b</td>
<td>Availability of cash point facilities</td>
<td>x2</td>
</tr>
<tr>
<td>Q08c</td>
<td>Availability of express checkout counters</td>
<td>x3</td>
</tr>
<tr>
<td>Q08d</td>
<td>Availability of parent and baby facilities</td>
<td>x4</td>
</tr>
<tr>
<td>Q08e</td>
<td>Availability of petrol station</td>
<td>x5</td>
</tr>
<tr>
<td>Q08f</td>
<td>Availability of restaurants or cafeteria</td>
<td>x6</td>
</tr>
<tr>
<td>Q08g</td>
<td>Availability of loyalty discount scheme</td>
<td>x7</td>
</tr>
<tr>
<td>Q08h</td>
<td>Car parking facilities</td>
<td>x8</td>
</tr>
<tr>
<td>Q08i</td>
<td>Convenient location</td>
<td>x9</td>
</tr>
<tr>
<td>Q08j</td>
<td>Customer service and assistance facilities</td>
<td>x10</td>
</tr>
<tr>
<td>Q08k</td>
<td>Frequency of special promotions</td>
<td>x11</td>
</tr>
<tr>
<td>Q08l</td>
<td>Friendliness of staff</td>
<td>x12</td>
</tr>
<tr>
<td>Q08m</td>
<td>General atmosphere of the store</td>
<td>x13</td>
</tr>
<tr>
<td>Q08n</td>
<td>Help with packing at checkout</td>
<td>x14</td>
</tr>
<tr>
<td>Q08o</td>
<td>High proportion of “own brand” products</td>
<td>x15</td>
</tr>
<tr>
<td>Q08p</td>
<td>High proportion of “value range” products</td>
<td>x16</td>
</tr>
<tr>
<td>Q08q</td>
<td>Length of queues at counters and checkouts</td>
<td>x17</td>
</tr>
<tr>
<td>Q08r</td>
<td>Low prices</td>
<td>x18</td>
</tr>
<tr>
<td>Q08s</td>
<td>Quality of fresh produce</td>
<td>x19</td>
</tr>
<tr>
<td>Q08t</td>
<td>Quality of packaged goods</td>
<td>x20</td>
</tr>
<tr>
<td>Q08u</td>
<td>Quality of trolleys</td>
<td>x21</td>
</tr>
<tr>
<td>Q08v</td>
<td>Transport provided</td>
<td>x22</td>
</tr>
</tbody>
</table>
4.2 Results and interpretation:

Data analysis process is initiated seeking to understand the relationship between the different domains of the Child Status Index. In the scatter matrix plots, figure 1, 2 and 3 below, the correlations between different domains are discussed. Here the scatter matrices have been set as to include domains that would be expected to have a high degree of association.

![Scatter matrix between care, education, shelter and social domains of Child Status Index](image)

**Figure 4.1 : Scatter matrix between care, education, shelter and social domains of Child Status Index**

The density plots that are observed on the diagonal of the scatter matrix suggest that there is a high association between the social and care sub-domains, and education and shelter sub-
domains. The scatter plots between shelter and social sub-domains indicate an almost horizontal relationship. This could be understood as suggest that a relationship is constant.

Figure 4.2: Scatter matrix between performance, wellness, abuse and emotional status

It is noted in figure 4.2 that wellness and abuse have almost similar densities but for the slight differences in the periods of the oscillatory plots. That is, wellness density plot seem to be slightly spread out compared to that of abuse.

Density plots of food and nutrition sub-domains indicate a very high association. Precisely this relationship suggests a correlation of 1 as seen figure 4.3 below.
Table 4.1 below provides the Pearson’s correlation of the sub-domains. All the relationships between the sub-domains are provided. Here it is noted the association observed in the scatter matrices above for food and nutrition is actually 1 as suggested. High positive associations are also noted between performance and education, emotional status and abuse, and social and emotional. These associations are more positive 0.5: 0.810, 0.535 and 0.562 respectively.

Food security and nutrition have very low positive associations with the other sub-domains. However, it is important to note that the two have exact similar associations with other sub-
domains. Measurement on either food security or nutrition would provide similar assessment of a child's wellbeing on this domain. This could be understood as to suggest that the measurer/respondent would obviously associate food security and nutrition highly. At this point it is noted that food security sub-domain as detailed in the Child Status Index focuses on the household and nutrition sub-domain on individual child. This therefore partly explains the source of subjectivity of the CSI as a measurement tool.

Some of the questions that are that need to be answered in the use of the Child Status Index include: 1. Does the tool apply to all ages?-Noting that infants and under-fives tend to show adverse symptoms of an ailment. 2. In the case of terminal illnesses like HIV and AIDS what would be the general notable differences in variation of scores?

An attempt to reduce the number of variables to be considered in the measurement of the child welfare status is conducted use the factor analysis. That is, to check whether the data collected in the 12 CSI sub-domains could be used to generate factor scores for the unobservable factors that affect the child welfare status say, HIV status, Age e.t.c.
Figure 4.4: A screen shot of factor analysis on Rcmdr package on R environment
Table 4.1: Pearson’s Correlation Coefficients between the sub-domains of Child Status Index

<table>
<thead>
<tr>
<th></th>
<th>ABUSE</th>
<th>CARE</th>
<th>EDUCATION</th>
<th>EMOTIONAL</th>
<th>FOOD</th>
<th>HEALTH</th>
<th>NUTRITION</th>
<th>PERFORMANCE</th>
<th>SHELTER</th>
<th>SOCIAL</th>
<th>WELLNESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABUSE</td>
<td>1.000</td>
<td>0.397</td>
<td>0.401</td>
<td>0.535</td>
<td>0.261</td>
<td>0.300</td>
<td>0.261</td>
<td>0.456</td>
<td>0.092</td>
<td>0.346</td>
<td>0.241</td>
</tr>
<tr>
<td>CARE</td>
<td>0.397</td>
<td>1.000</td>
<td>0.312</td>
<td>0.426</td>
<td>0.360</td>
<td>0.440</td>
<td>0.360</td>
<td>0.401</td>
<td>0.225</td>
<td>0.393</td>
<td>0.227</td>
</tr>
<tr>
<td>EDUCATION</td>
<td>0.401</td>
<td>0.312</td>
<td>1.000</td>
<td>0.355</td>
<td>0.240</td>
<td>0.256</td>
<td>0.240</td>
<td>0.809</td>
<td>0.017</td>
<td>0.315</td>
<td>0.262</td>
</tr>
<tr>
<td>EMOTIONAL</td>
<td>0.535</td>
<td>0.426</td>
<td>0.355</td>
<td>1.000</td>
<td>0.280</td>
<td>0.371</td>
<td>0.279</td>
<td>0.449</td>
<td>0.084</td>
<td>0.562</td>
<td>0.428</td>
</tr>
<tr>
<td>FOOD</td>
<td>0.261</td>
<td>0.360</td>
<td>0.230</td>
<td>0.280</td>
<td>1.000</td>
<td>0.366</td>
<td>1.000</td>
<td>0.338</td>
<td>0.334</td>
<td>0.254</td>
<td>0.422</td>
</tr>
<tr>
<td>HEALTH</td>
<td>0.297</td>
<td>0.440</td>
<td>0.256</td>
<td>0.371</td>
<td>0.366</td>
<td>1.000</td>
<td>0.366</td>
<td>0.297</td>
<td>0.157</td>
<td>0.203</td>
<td>0.329</td>
</tr>
<tr>
<td>NUTRITION</td>
<td>0.261</td>
<td>0.360</td>
<td>0.240</td>
<td>0.280</td>
<td>1.000</td>
<td>0.366</td>
<td>1.000</td>
<td>0.338</td>
<td>0.334</td>
<td>0.254</td>
<td>0.422</td>
</tr>
<tr>
<td>PERFORMANCE</td>
<td>0.456</td>
<td>0.400</td>
<td>0.809</td>
<td>0.450</td>
<td>0.337</td>
<td>0.297</td>
<td>0.338</td>
<td>1.000</td>
<td>0.090</td>
<td>0.373</td>
<td>0.260</td>
</tr>
<tr>
<td>SHELTER</td>
<td>0.092</td>
<td>0.224</td>
<td>0.017</td>
<td>0.084</td>
<td>0.334</td>
<td>0.157</td>
<td>0.334</td>
<td>0.090</td>
<td>1.000</td>
<td>0.005</td>
<td>0.154</td>
</tr>
<tr>
<td>SOCIAL</td>
<td>0.346</td>
<td>0.393</td>
<td>0.315</td>
<td>0.562</td>
<td>0.254</td>
<td>0.202</td>
<td>0.254</td>
<td>0.373</td>
<td>0.005</td>
<td>1.000</td>
<td>0.272</td>
</tr>
<tr>
<td>WELLNESS</td>
<td>0.241</td>
<td>0.227</td>
<td>0.262</td>
<td>0.428</td>
<td>0.422</td>
<td>0.329</td>
<td>0.422</td>
<td>0.260</td>
<td>0.154</td>
<td>0.273</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4.2: Principal components (proportions of variance): Unreduced

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
<th>C16</th>
<th>C17</th>
<th>C18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sd</td>
<td>2.21</td>
<td>1.41</td>
<td>1.24</td>
<td>1.12</td>
<td>1.04</td>
<td>1.02</td>
<td>0.90</td>
<td>0.86</td>
<td>0.84</td>
<td>0.80</td>
<td>0.77</td>
<td>0.75</td>
<td>0.71</td>
<td>0.66</td>
<td>0.61</td>
<td>0.55</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td>Prop</td>
<td>0.27</td>
<td>0.11</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Cum</td>
<td>0.27</td>
<td>0.38</td>
<td>0.47</td>
<td>0.54</td>
<td>0.60</td>
<td>0.66</td>
<td>0.70</td>
<td>0.74</td>
<td>0.78</td>
<td>0.82</td>
<td>0.85</td>
<td>0.88</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
<td>0.97</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sd - standard deviation, Prop – proportion of variance explained by the component and Cum – cumulative variance by components
Using the Rcommander package of R software, R.2.15.1, we attempt to conduct factor analysis.

A screen print of the how this is done is provided below in figure 4. This error output is returned:

```r
> .FA <-
+ factanal(~ABUSE+CARE+EDUCATION+EMOTIONAL+FOOD.SECURITY+H..CARE+NUTRITION+PERFORMANCE+SHELTER+SOCIAL+WELLNESS,
+ factors=6, rotation="none", scores="none", data=data)
> .FA
[1] "Error in solve.default(cv) : 
Lapack routine dgesv: system is exactly singular"
attr(,"class")
[1] "try-error"
attr(,"condition")
<simpleError in solve.default(cv): Lapack routine dgesv: system is exactly singular>
```

This output states that the inverse matrix of the correlation matrix provide in table 4.2 does not exists. At this point we note that it is therefore not possible to obtain the unobserved factors for the child welfare status starting with the sample correlation matrix. Given the noted associations earlier between sub-domains one would conclude that they have a part in the observed singularity of the system. To deal with this challenge one of the variables is dropped for the pairs that have correlations tending towards 1. In this case dropping these variables reduces the measured variables to be considered for factor analysis. It is therefore not valuable to proceed with the reduction of variables by factor analysis.

Measurement of child welfare could be improved by use of a Likert’s 5-point scale (not discussed in this work) on questions that would be set so as to best collect data. The collected
data could then be analyzed as detailed in the remaining sections of this chapter to obtain the unobserved factors for child welfare status. The obtained factors are obvious. The naming of these factors will depend on the knowledge of the researcher on the subject. The number of the factors to be fitted in the final model is also a choice that the researcher would need to make. However, though obvious factors will be eventually fitted, factor analysis allows a quantification of these factors. That is, we obtained factor scores.

To follow through the discussion on how factor analysis could be conducted on a set of ordinal data we now consider a data set of ordinal variables. The interest is to discuss the process of identifying unobserved factors, corresponding factors loadings and weights that are generated simultaneously to avoid a situation where they would be correlated.
A summary of the dataset is provided in the table 4.4 below:

**Table 4.3: Summary of dataset used in discussing factor analysis process**

<table>
<thead>
<tr>
<th>Measured Variables</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>n</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>2.5900</td>
<td>1.649</td>
<td>517</td>
<td>29</td>
</tr>
<tr>
<td>x3</td>
<td>3.335</td>
<td>1.535</td>
<td>517</td>
<td>29</td>
</tr>
<tr>
<td>x5</td>
<td>2.990</td>
<td>1.670</td>
<td>519</td>
<td>27</td>
</tr>
<tr>
<td>x6</td>
<td>2.314</td>
<td>1.474</td>
<td>519</td>
<td>27</td>
</tr>
<tr>
<td>x7</td>
<td>3.285</td>
<td>1.514</td>
<td>519</td>
<td>27</td>
</tr>
<tr>
<td>x8</td>
<td>3.859</td>
<td>1.620</td>
<td>517</td>
<td>29</td>
</tr>
<tr>
<td>x9</td>
<td>4.338</td>
<td>1.002</td>
<td>517</td>
<td>29</td>
</tr>
<tr>
<td>x10</td>
<td>3.953</td>
<td>1.184</td>
<td>515</td>
<td>31</td>
</tr>
<tr>
<td>x11</td>
<td>3.944</td>
<td>1.232</td>
<td>517</td>
<td>29</td>
</tr>
<tr>
<td>x12</td>
<td>4.148</td>
<td>1.095</td>
<td>527</td>
<td>19</td>
</tr>
<tr>
<td>x13</td>
<td>3.990</td>
<td>1.122</td>
<td>524</td>
<td>22</td>
</tr>
<tr>
<td>x15</td>
<td>3.078</td>
<td>1.476</td>
<td>523</td>
<td>23</td>
</tr>
<tr>
<td>x16</td>
<td>3.524</td>
<td>1.391</td>
<td>515</td>
<td>31</td>
</tr>
<tr>
<td>x17</td>
<td>4.369</td>
<td>0.990</td>
<td>520</td>
<td>26</td>
</tr>
<tr>
<td>x18</td>
<td>4.507</td>
<td>0.905</td>
<td>521</td>
<td>25</td>
</tr>
<tr>
<td>x19</td>
<td>4.556</td>
<td>0.920</td>
<td>522</td>
<td>24</td>
</tr>
<tr>
<td>x21</td>
<td>3.784</td>
<td>1.238</td>
<td>518</td>
<td>28</td>
</tr>
<tr>
<td>x22</td>
<td>1.936</td>
<td>1.457</td>
<td>503</td>
<td>43</td>
</tr>
</tbody>
</table>
An exploration of the associations between different sets of measured variables is provided in figure 4.5, 4.6 and 4.7.

Figure 4.5: Scatter matrix plot between variables $x_2$, $x_3$, $x_5$, $x_6$ and $x_7$

The density plots for variables $x_2$ and $x_6$ look similar, $x_3$ and $x_7$ too.
Figure 4.6: Scatter matrix plot between variables $x_7$, $x_8$, $x_9$, $x_{10}$, $x_{11}$ and $x_{12}$
The density plots for $x_{17}$ and $x_{18}$ look similar.
Figure 4.8: Scatter matrix plot between variables x19, x21 and x22

Now, the iterative process of factor analysis requires obtain the first estimate of the factor loadings. Then repeat the procedure of estimating the communalities until there is convergence. This estimate is obtained by using the principal components (eigen values) and corresponding weights (eigen vectors). Note that principal components explain 100% variation in the measured variables (see table 2 above). The first six components of 18 explain 59.7% of the variation in the data. A good estimate therefore is to start by fitting a 5-factor model.
The scree plot in figure 4.9 below suggests that the best choice of factors would be between factors 5 and 10.

Figure 4.9: Scree plot of the principal components

Rcommander allows a maximum of 12 factors to be fitted on the data. Table 4.5 and 4.6 below provide the factor loadings for un-rotated solution and the corresponding variances.
Table 4.4: Factor loadings for a 12 factors fitted model (un-rotated)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>UNIQUENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>0.350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.593</td>
</tr>
<tr>
<td>x3</td>
<td>0.334</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.440</td>
</tr>
<tr>
<td>x5</td>
<td>0.584</td>
<td>-0.546</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>x6</td>
<td>0.286</td>
<td>-0.164</td>
<td>0.118</td>
<td>0.225</td>
<td>0.307</td>
<td></td>
<td>-0.112</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.638</td>
</tr>
<tr>
<td>x7</td>
<td>0.606</td>
<td></td>
<td></td>
<td>0.640</td>
<td>0.465</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>x8</td>
<td>0.462</td>
<td>-0.312</td>
<td>0.151</td>
<td>-0.108</td>
<td>-0.312</td>
<td>0.169</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.489</td>
</tr>
<tr>
<td>x9</td>
<td>0.271</td>
<td>0.241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.346</td>
<td></td>
<td></td>
<td></td>
<td>-0.271</td>
<td>0.123</td>
<td>0.575</td>
</tr>
<tr>
<td>x10</td>
<td>0.566</td>
<td>-0.110</td>
<td>0.453</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.278</td>
<td>0.137</td>
<td>0.229</td>
<td>-0.212</td>
<td>-0.287</td>
<td></td>
</tr>
<tr>
<td>x11</td>
<td>0.465</td>
<td>0.517</td>
<td>0.127</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.113</td>
<td>0.123</td>
<td>-0.106</td>
<td></td>
<td></td>
<td>-0.152</td>
</tr>
<tr>
<td>x12</td>
<td>0.627</td>
<td>0.306</td>
<td>-0.200</td>
<td>0.163</td>
<td>0.222</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.356</td>
</tr>
<tr>
<td>x13</td>
<td>0.725</td>
<td></td>
<td>-0.425</td>
<td>0.142</td>
<td>0.518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>x15</td>
<td>0.322</td>
<td>0.851</td>
<td>0.215</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>x16</td>
<td>0.220</td>
<td>0.636</td>
<td>0.115</td>
<td></td>
<td></td>
<td></td>
<td>-0.103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.413</td>
</tr>
<tr>
<td>x17</td>
<td>0.443</td>
<td>0.404</td>
<td>0.110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.582</td>
</tr>
<tr>
<td>x18</td>
<td>0.305</td>
<td>0.150</td>
<td>0.616</td>
<td></td>
<td></td>
<td></td>
<td>-0.273</td>
<td>0.122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.298</td>
</tr>
<tr>
<td>x19</td>
<td>0.408</td>
<td>0.226</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.138</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.545</td>
</tr>
<tr>
<td>x21</td>
<td>0.716</td>
<td></td>
<td>-0.433</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>x22</td>
<td>0.199</td>
<td>0.205</td>
<td>-0.110</td>
<td>0.631</td>
<td>0.113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.331</td>
</tr>
</tbody>
</table>

Table 4.5: Proportion and cumulative variances for the 12 fitted factor

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>SS LOADINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.927</td>
<td>1.669</td>
<td>1.298</td>
<td>1.065</td>
<td>0.896</td>
<td>0.882</td>
<td>0.658</td>
<td>0.497</td>
<td>0.430</td>
<td>0.355</td>
<td>0.273</td>
<td>0.166</td>
<td>0.827</td>
</tr>
<tr>
<td>PROPORTION VAR</td>
<td>0.218</td>
<td>0.093</td>
<td>0.072</td>
<td>0.059</td>
<td>0.050</td>
<td>0.049</td>
<td>0.037</td>
<td>0.028</td>
<td>0.024</td>
<td>0.020</td>
<td>0.015</td>
<td>0.009</td>
<td>0.218</td>
</tr>
<tr>
<td>CUMULATIVE VAR</td>
<td>0.218</td>
<td>0.311</td>
<td>0.383</td>
<td>0.442</td>
<td>0.492</td>
<td>0.541</td>
<td>0.578</td>
<td>0.605</td>
<td>0.629</td>
<td>0.649</td>
<td>0.664</td>
<td>0.673</td>
<td>0.218</td>
</tr>
</tbody>
</table>
Table 4.6: Factor loadings for a 12 factors fitted model (rotated)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>UNIQUENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>0.295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.116</td>
<td>0.384</td>
<td>0.167</td>
<td>0.114</td>
<td></td>
<td></td>
<td>0.330</td>
<td>0.593</td>
</tr>
<tr>
<td>x3</td>
<td></td>
<td>0.107</td>
<td>0.125</td>
<td>0.111</td>
<td>0.116</td>
<td></td>
<td>0.684</td>
<td>0.104</td>
<td>0.139</td>
<td></td>
<td></td>
<td></td>
<td>0.440</td>
</tr>
<tr>
<td>x5</td>
<td>0.961</td>
<td></td>
<td>0.110</td>
<td></td>
<td></td>
<td></td>
<td>0.115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.136</td>
</tr>
<tr>
<td>x6</td>
<td>0.194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.225</td>
<td>0.228</td>
<td>0.349</td>
<td>0.220</td>
<td></td>
<td>-0.155</td>
<td></td>
<td>0.638</td>
</tr>
<tr>
<td>x7</td>
<td>0.165</td>
<td></td>
<td>0.106</td>
<td>0.131</td>
<td>0.933</td>
<td></td>
<td>0.134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x8</td>
<td>0.636</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.113</td>
<td></td>
<td></td>
<td>-0.108</td>
<td>0.161</td>
<td>0.172</td>
<td>0.489</td>
</tr>
<tr>
<td>x9</td>
<td>0.103</td>
<td></td>
<td>0.110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x10</td>
<td>0.160</td>
<td>0.285</td>
<td></td>
<td></td>
<td></td>
<td>0.209</td>
<td>0.166</td>
<td>0.130</td>
<td>0.157</td>
<td>0.589</td>
<td></td>
<td>0.103</td>
<td>0.382</td>
</tr>
<tr>
<td>x11</td>
<td>0.224</td>
<td></td>
<td>0.485</td>
<td>0.225</td>
<td></td>
<td></td>
<td>0.325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.118</td>
</tr>
<tr>
<td>x12</td>
<td>0.113</td>
<td>0.563</td>
<td></td>
<td></td>
<td></td>
<td>0.297</td>
<td>0.125</td>
<td>0.177</td>
<td>0.105</td>
<td>0.336</td>
<td></td>
<td>0.107</td>
<td>0.200</td>
</tr>
<tr>
<td>x13</td>
<td>0.121</td>
<td>0.941</td>
<td></td>
<td></td>
<td></td>
<td>0.120</td>
<td></td>
<td>0.142</td>
<td>0.103</td>
<td></td>
<td></td>
<td></td>
<td>0.150</td>
</tr>
<tr>
<td>x15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.969</td>
<td></td>
<td></td>
<td>0.105</td>
<td></td>
<td>0.105</td>
</tr>
<tr>
<td>x16</td>
<td>-0.107</td>
<td></td>
<td>0.699</td>
<td>0.217</td>
<td></td>
<td></td>
<td>0.268</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x17</td>
<td>0.148</td>
<td>0.211</td>
<td></td>
<td></td>
<td></td>
<td>0.378</td>
<td>0.125</td>
<td>0.143</td>
<td>0.216</td>
<td></td>
<td>0.201</td>
<td></td>
<td>0.582</td>
</tr>
<tr>
<td>x18</td>
<td>0.114</td>
<td>0.170</td>
<td>0.785</td>
<td></td>
<td>0.104</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.268</td>
<td></td>
<td>0.131</td>
<td>0.298</td>
</tr>
<tr>
<td>x19</td>
<td>0.213</td>
<td>0.205</td>
<td>0.231</td>
<td></td>
<td>0.131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.524</td>
<td></td>
<td>0.545</td>
</tr>
<tr>
<td>x21</td>
<td>0.150</td>
<td>0.221</td>
<td></td>
<td>0.148</td>
<td>0.919</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.116</td>
<td>0.124</td>
</tr>
<tr>
<td>x22</td>
<td>0.115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.784</td>
</tr>
</tbody>
</table>

Table 4.7: Proportion and cumulative variances of 12-factor loadings (rotated)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS LOADINGS</td>
<td>1.667</td>
<td>1.543</td>
<td>1.533</td>
<td>1.310</td>
<td>1.080</td>
<td>0.995</td>
<td>0.820</td>
<td>0.810</td>
<td>0.766</td>
<td>0.754</td>
<td>0.472</td>
<td>0.367</td>
</tr>
<tr>
<td>PROPORTION VAR</td>
<td>0.093</td>
<td>0.086</td>
<td>0.085</td>
<td>0.073</td>
<td>0.060</td>
<td>0.055</td>
<td>0.046</td>
<td>0.045</td>
<td>0.043</td>
<td>0.042</td>
<td>0.026</td>
<td>0.020</td>
</tr>
<tr>
<td>CUMULATIVE VAR</td>
<td>0.093</td>
<td>0.178</td>
<td>0.263</td>
<td>0.336</td>
<td>0.396</td>
<td>0.452</td>
<td>0.497</td>
<td>0.542</td>
<td>0.585</td>
<td>0.627</td>
<td>0.653</td>
<td>0.673</td>
</tr>
</tbody>
</table>
In table 4.5 and 4.6 we note that most of the variation in the measured variables is explained by a few factors. F1 is highlighted in red in table 4.6 and 4.8. The proportion of variation explained by this factor is 21.8% and 9.3% respectively for the un-rotated and rotated solution of factor loadings. Using Varimax in Rcommander orthogonal rotation is conducted. The rotation does not provide a better fit, instead ensures more factors are included by distributing variation. After the rotation more factors tend to have values tending to 1 for different variables. Under the uniqueness column for both results values tending towards 1 are highlighted in blue. These values suggest that less of the variation in the corresponding measured variable is explained by the set of factors. Here it is noted that the smaller the value of uniqueness the better a fit of the factors on the measured variables.

Also provided in the results is information on goodness of fit. The following is the result:

Test of the hypothesis that 12 factors are sufficient.

The chi square statistic is 1.56 on 3 degrees of freedom.

The p-value is 0.669

Using the p-value we can conclude that for a 12-factor model to sufficiently explain the variation, a high value of the error must be allowable. Precisely an error equal to or greater than 66.9% should be allowable. This result is same for the 12-factor loadings fit un-rotated and rotated.

Going back to the first choice of fitting 5 factors, table 4.9 and 4.10 below provide the factor loadings and the corresponding variances respectively. The 5 factors explain 46.3% of the total variation. Considering the entries of the uniqueness column it is noted that 9 measured variables have a put fit of the factors. That is, they have values of uniqueness tending towards 1. This is 4
Compare to the result discussed earlier of a 12-factor loadings model. This implies that less will be associated with the factors. Therefore, a solution that explains a huge proportion of variance, has most values of the uniqueness is tending towards zero and sufficiently fits the given the allowable error is desired.

Table 4.8: Factor loadings for a 5 factors fitted model (rotated)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>UNIQUENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>0.462</td>
<td></td>
<td></td>
<td>0.264</td>
<td></td>
<td>0.703</td>
</tr>
<tr>
<td>x3</td>
<td>0.159</td>
<td>0.410</td>
<td>0.145</td>
<td></td>
<td></td>
<td>0.756</td>
</tr>
<tr>
<td>x5</td>
<td>0.180</td>
<td>0.280</td>
<td></td>
<td>0.764</td>
<td>-0.132</td>
<td>0.284</td>
</tr>
<tr>
<td>x6</td>
<td>0.510</td>
<td>0.138</td>
<td>0.115</td>
<td>-0.115</td>
<td></td>
<td>0.690</td>
</tr>
<tr>
<td>x7</td>
<td>0.120</td>
<td>0.452</td>
<td>0.322</td>
<td>0.183</td>
<td>0.107</td>
<td>0.633</td>
</tr>
<tr>
<td>x8</td>
<td>0.181</td>
<td></td>
<td>0.778</td>
<td></td>
<td></td>
<td>0.353</td>
</tr>
<tr>
<td>x9</td>
<td>0.236</td>
<td>0.118</td>
<td>0.193</td>
<td>0.134</td>
<td></td>
<td>0.875</td>
</tr>
<tr>
<td>x10</td>
<td>0.457</td>
<td>0.432</td>
<td>0.387</td>
<td>0.144</td>
<td></td>
<td>0.432</td>
</tr>
<tr>
<td>x11</td>
<td>0.272</td>
<td>0.371</td>
<td>0.543</td>
<td></td>
<td></td>
<td>0.489</td>
</tr>
<tr>
<td>x12</td>
<td>0.728</td>
<td>0.209</td>
<td>0.335</td>
<td></td>
<td></td>
<td>0.309</td>
</tr>
<tr>
<td>x13</td>
<td>0.824</td>
<td>0.162</td>
<td></td>
<td></td>
<td></td>
<td>0.278</td>
</tr>
<tr>
<td>x15</td>
<td>0.161</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.815</td>
</tr>
<tr>
<td>x16</td>
<td>0.170 &amp;-0.112</td>
<td>0.793</td>
<td>0.324</td>
<td></td>
<td></td>
<td>0.308</td>
</tr>
<tr>
<td>x17</td>
<td>0.361</td>
<td>0.187</td>
<td>0.455</td>
<td>0.137</td>
<td></td>
<td>0.609</td>
</tr>
<tr>
<td>x18</td>
<td>0.232</td>
<td>0.674</td>
<td></td>
<td>0.204</td>
<td>0.445</td>
<td></td>
</tr>
<tr>
<td>x19</td>
<td>0.367</td>
<td></td>
<td>0.276</td>
<td>0.247</td>
<td></td>
<td>0.728</td>
</tr>
<tr>
<td>x21</td>
<td>0.470</td>
<td>0.126</td>
<td>0.140</td>
<td>0.169</td>
<td>0.150</td>
<td>0.693</td>
</tr>
<tr>
<td>x22</td>
<td>0.103</td>
<td>0.402</td>
<td>0.104</td>
<td>-0.183</td>
<td>0.164</td>
<td>0.757</td>
</tr>
</tbody>
</table>

Table 4.9: Proportion and cumulative variances of 5-factor loadings fitted

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS LOADINGS</td>
<td>2.225</td>
<td>1.578</td>
<td>1.553</td>
<td>1.516</td>
<td>1.463</td>
</tr>
<tr>
<td>PROPORTION VAR</td>
<td>0.124</td>
<td>0.088</td>
<td>0.086</td>
<td>0.084</td>
<td>0.081</td>
</tr>
<tr>
<td>CUMULATIVE VAR</td>
<td>0.124</td>
<td>0.211</td>
<td>0.298</td>
<td>0.382</td>
<td>0.463</td>
</tr>
</tbody>
</table>
the following result is outputted.

Of the hypothesis that 5 factors are sufficient.

The chi square statistic is 199.78 on 73 degrees of freedom.

The p-value is 9.95e-14

A sensible error can be set at value very low given the p-value.

As a result, the factors model is fitted. The results are provided in Table 4.11 and 4.12 below.

| Table 4.10: Factor loadings for a 7-factors fitted model (rotated) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| F1   | F2   | F3   | F4   | F5   | F6   | F7   | UNIQUENESS |
| 0.274 | 0.337 |   | 0.118 | 0.253 | 0.722 |
| 0.120 | 0.336 | 0.162 | 0.223 | 0.170 | 0.758 |
| 0.161 | 0.281 | 0.846 | -0.115 | 0.164 |
| 0.564 | 0.140 | -0.100 | -0.114 | 0.180 | 0.600 |
| 0.141 | 0.521 | 0.166 | 0.123 | 0.141 | 0.637 |
| 0.171 | 0.700 | 0.158 | 0.443 |
| 0.168 | 0.114 | 0.579 | 0.615 |
| 0.445 | 0.497 | 0.130 | 0.397 | 0.367 |
| 0.295 | 0.549 | 0.278 | 0.156 | 0.496 |
| 0.820 | 0.336 | 0.141 | 0.390 |
| 0.735 | 0.154 | 0.107 | 0.921 | 0.115 | 0.127 |
|       |       |       |       | 0.709 | 0.156 | 0.445 |
| 0.365 | 0.287 | 0.121 | 0.244 | 0.313 | 0.611 |
| 0.261 | 0.158 | 0.177 | 0.925 | 0.005 |
| 0.389 | 0.235 | 0.190 | 0.744 |
| 0.480 | 0.213 | 0.116 | 0.155 | 0.156 | 0.297 | 0.575 |
| 0.194 | -0.102 | 0.119 | 0.616 | 0.551 |

| Table 4.11: Proportion and cumulative variances of 7-factor loadings fitted |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| SS LOADINGS | F1   | F2   | F3   | F4   | F5   | F6   | F7   |
| 2.232  | 1.709 | 1.565 | 1.494 | 1.109 | 0.808 | 0.653 |
| PROPORTION VAR | 0.124 | 0.095 | 0.087 | 0.083 | 0.062 | 0.045 | 0.653 |
| CUMULATIVE VAR | 0.124 | 0.219 | 0.306 | 0.389 | 0.451 | 0.495 | 0.532 |
In regard to the explained variation this result is better than the 5-factor one. A proportion of 53.2% of the total variation is explained by the factors. However, it is important to note that the number of the value of uniqueness that is undesired still remains at 9. To investigate whether better results will be obtained with more factors, an 8-factor model is fitted. The results of these fit are provided in table 4.13 and 4.14 below.

Table 4.12: Factor loadings for an 8-factors fitted model (rotated)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>UNIQUENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>0.155</td>
<td></td>
<td>0.177</td>
<td></td>
<td>0.921</td>
<td>0.104</td>
<td>0.112</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>x3</td>
<td>0.133</td>
<td>0.290</td>
<td>0.169</td>
<td></td>
<td>0.246</td>
<td>0.216</td>
<td></td>
<td>0.748</td>
<td></td>
</tr>
<tr>
<td>x5</td>
<td>0.165</td>
<td>0.285</td>
<td>-0.114</td>
<td>0.737</td>
<td>0.211</td>
<td></td>
<td></td>
<td>0.284</td>
<td></td>
</tr>
<tr>
<td>x6</td>
<td>0.576</td>
<td></td>
<td>0.123</td>
<td></td>
<td>-0.102</td>
<td>0.170</td>
<td>0.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x7</td>
<td>0.134</td>
<td>0.552</td>
<td>0.150</td>
<td>0.182</td>
<td></td>
<td>0.151</td>
<td></td>
<td>0.588</td>
<td></td>
</tr>
<tr>
<td>x8</td>
<td>0.148</td>
<td></td>
<td>0.783</td>
<td></td>
<td>0.150</td>
<td></td>
<td>0.328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x9</td>
<td>0.166</td>
<td></td>
<td>0.100</td>
<td></td>
<td></td>
<td>0.583</td>
<td>0.611</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x10</td>
<td>0.450</td>
<td>0.478</td>
<td>0.126</td>
<td></td>
<td></td>
<td>0.419</td>
<td></td>
<td>0.358</td>
<td></td>
</tr>
<tr>
<td>x11</td>
<td>0.308</td>
<td>0.520</td>
<td></td>
<td>0.288</td>
<td>0.142</td>
<td>0.171</td>
<td></td>
<td>0.487</td>
<td></td>
</tr>
<tr>
<td>x12</td>
<td>0.817</td>
<td>0.309</td>
<td></td>
<td>0.118</td>
<td></td>
<td></td>
<td>0.204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x13</td>
<td>0.749</td>
<td>0.131</td>
<td></td>
<td></td>
<td>0.142</td>
<td></td>
<td>0.376</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x15</td>
<td></td>
<td></td>
<td></td>
<td>0.836</td>
<td></td>
<td></td>
<td></td>
<td>0.276</td>
<td></td>
</tr>
<tr>
<td>x16</td>
<td></td>
<td>0.780</td>
<td></td>
<td>-0.101</td>
<td>0.145</td>
<td></td>
<td></td>
<td>0.345</td>
<td></td>
</tr>
<tr>
<td>x17</td>
<td>0.360</td>
<td>0.278</td>
<td></td>
<td>0.137</td>
<td>0.248</td>
<td>0.325</td>
<td></td>
<td>0.604</td>
<td></td>
</tr>
<tr>
<td>x18</td>
<td>0.250</td>
<td>0.138</td>
<td>0.177</td>
<td></td>
<td>0.932</td>
<td></td>
<td></td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>x19</td>
<td>0.379</td>
<td></td>
<td>0.275</td>
<td>0.194</td>
<td></td>
<td></td>
<td>0.723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x21</td>
<td>0.475</td>
<td>0.120</td>
<td>0.218</td>
<td>0.153</td>
<td>0.119</td>
<td>0.139</td>
<td>0.265</td>
<td>0.585</td>
<td></td>
</tr>
<tr>
<td>x22</td>
<td>0.189</td>
<td>0.129</td>
<td>-0.120</td>
<td></td>
<td></td>
<td></td>
<td>0.630</td>
<td>0.517</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.13: Proportion and cumulative variances of 8-factor loadings fitted

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS LOADINGS</td>
<td>2.225</td>
<td>1.592</td>
<td>1.458</td>
<td>1.443</td>
<td>1.131</td>
<td>1.039</td>
<td>0.835</td>
<td>0.581</td>
</tr>
<tr>
<td>PROPORTION VAR</td>
<td>0.124</td>
<td>0.088</td>
<td>0.081</td>
<td>0.080</td>
<td>0.063</td>
<td>0.058</td>
<td>0.046</td>
<td>0.032</td>
</tr>
<tr>
<td>CUMULATIVE VAR</td>
<td>0.124</td>
<td>0.212</td>
<td>0.293</td>
<td>0.373</td>
<td>0.436</td>
<td>0.494</td>
<td>0.540</td>
<td>0.572</td>
</tr>
</tbody>
</table>

44
Given the result of the 8-factor fitted model we can conclude that this provides the best model insofar as the decision is made on a combination of variation of variance explained and values of the uniqueness. Considering the figures highlighted in blue in Table 4.13 we note there 8 values more than 0.5, 1 less the results obtained for a 5 and 7-factor model fit. This means that one more manifest variable (referred mostly in this work as measured variable) provides useful information that can be utilized in determining the factor names/labels. Table 4.14 shows that the fitted factors explain 57.2% of the variation. Factor loadings highlighted in red are of values greater than 0.4 and provide information on how to link the measured variables to the generated factors. Figure 4.10 below provides a pictorial presentation of the linkages.

Variables x3, x17 and x19 highlighted in red are not linked to any of the factors. They have low values to determine any associations. Calculating communalities with these variables included yields an average of 0.572 is equivalent to 57.2% explained variance. The result therefore can be improved by dropping the variables by 0.053, that is, 62.5% variation explained.

Finally, we have a result that appropriately fits a factor model and we can express one of the measured variables, say x21 as follows:

\[ x_{21} = 3.784 + 0.475F_1 + 0.120F_3 + 0.218F_4 + 0.153F_5 + 0.119F_6 + 0.139F_7 + 0.265F_8 + 0.585 \]

A generation and quantification of the underlying factors through the process of iterative factor analysis is complete for a set ordinal data.
Figure 4.10: Measured variables linked with factors generated
Chapter 5: Conclusions and Recommendations

Starting with results of principal components analysis the study showed how iterative principal factor analysis is conducted to reduce the number of factors from 18 to 8. Simultaneous generation of factors loadings and scores is discussed. Eventually an 8-factor model is fitted to a set of ordinal data. The factors together explain 62.5% and sufficiently fit the data with an allowable error of not less than 9.88%. The factors obtained are quantified and provide useful unobserved variables that facilitate better understanding of the relationships between observed variables. Such a procedure can be useful in definition of an index that measures attributes of subject of interest.

An improvement of child welfare status tool could be obtained if a set of questions with responses defined on 5-point ordinal scale is designed making it possible to use IPF Analysis for generation of the main factors (unobservable) and therefore dealing with errors due to quality of the measurement tool.
Index

A: R Script

data=read.csv("CSI_data.csv", header=T)
attach(data)#ensures that we use the correct dataset
summary(data)#summarises the CSI_data
hetcor(data)# generates the pearson's, polychoric and polyserials correlations
detach(data)
library(abind, pos=4)
> .FA <-
  + factanal(~ABUSE+CARE+EDUCATION+EMOTIONAL+FOOD.SECURITY+H..CARE+NUTRITION+PERFORMANCE+SHELTER+SOCIAL+WELLNESS,
  +  factors=6, rotation="none", scores="none", data=data)
> .FA

[1] "Error in solve.default(cv) : 
  Lapack routine dgesv: system is exactly singular
"
attr(, "class")
[1] "try-error"
attr(,"condition")

<simpleError in solve.default(cv): Lapack routine dgesv: system is exactly singular>
> remove(.FA)

> Hist(data$WELLNESS, scale="frequency", breaks="Sturges", col="darkgray")
>
> scatterplotMatrix(~ABUSE+CARE+EDUCATION+EMOTIONAL+FOOD.SECURITY+H..CARE+NUTRITION+PERFORMANCE+SHELTER+SOCIAL+WELLNESS,
  +  reg.line=lm, smooth=TRUE, spread=TRUE, span=0.5, diagonal = 'density', +  data=data)
> scatterplotMatrix(~ABUSE+CARE+EDUCATION+EMOTIONAL+FOOD.SECURITY,
  +  reg.line=lm, smooth=TRUE, spread=TRUE, span=0.5, diagonal = 'density', +  data=data)
> scatterplotMatrix(~ABUSE+CARE+EDUCATION+EMOTIONAL+FOOD.SECURITY, 
+ reg.line=lm, smooth=TRUE, spread=TRUE, span=0.5, diagonal = 'histogram', 
+ data=data) 

> scatterplotMatrix(~ABUSE+CARE+EDUCATION+EMOTIONAL+FOOD.SECURITY, 
+ reg.line=lm, smooth=TRUE, spread=TRUE, span=0.5, diagonal = 'density', 
+ data=data) 

> scatterplotMatrix(~CARE+EDUCATION+SHELTER+SOCIAL, reg.line=lm, smooth=TRUE, 
+ spread=TRUE, span=0.5, diagonal = 'density', data=data) 

> scatterplotMatrix(~ABUSE+EMOTIONAL+PERFORMANCE+WELLNESS, reg.line=lm, 
+ smooth=TRUE, spread=TRUE, span=0.5, diagonal = 'density', data=data) 

> scatterplotMatrix(~FOOD.SECURITY+H..CARE+NUTRITION, reg.line=lm, 
+ smooth=TRUE, spread=TRUE, span=0.5, diagonal = 'density', data=data) 

> cor(data[,c("ABUSE","CARE","EDUCATION","EMOTIONAL","FOOD.SECURITY", 
+ "H..CARE","NUTRITION","PERFORMANCE","SHELTER","SOCIAL","WELLNESS")], 
+ use="complete.obs") 
hetcor(data)# generates the detach(data)pearson's, polychoric and polyserials correlations 

. PC <- 
+ princomp(~x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22, 
+ cor=TRUE, data=Factor_new1) 
screeplot(.PC) 
> remove(.PC)
> .FA <-
+ factanal(~x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22,
+ factors=12, rotation="none", scores="none", data=Factor_new1)

> .FA
Call:
factanal(x = ~x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x15 + x16 + x17 + x18 + x19 + x21 + x22, factors = 12, data = Factor_new1, scores = "none", rotation = "none")

> .FA <-
+ factanal(~x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22,
+ factors=12, rotation="varimax", scores="none", data=Factor_new1)

> .FA
Call:
factanal(x = ~x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x15 + x16 + x17 + x18 + x19 + x21 + x22, factors = 12, data = Factor_new1, scores = "none", rotation = "varimax")

remove(.FA)

> .FA <-
+ factanal(~x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22,
+ factors=12, rotation="varimax", scores="regression", data=Factor_new1)

> .FA
Call:
factanal(x = ~x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x15 + x16 + x17 + x18 + x19 + x21 + x22, factors = 12, data = Factor_new1, scores = "regression", rotation = "varimax")

FA <-
+ factanal(~x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22,
+ factors=5, rotation="varimax", scores="regression", data=Factor_new1)
> FA

Call:

`factanal(x = -x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x15 + x16 + x17 + x18 + x19 + x21 + x22, factors = 5, data = Factor_new1, scores = "regression", rotation = "varimax")`

FA <-

- `factanal(-x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22, factors=8, rotation="varimax", scores="regression", data=Factor_new1)`

> FA

Call:

`factanal(x = -x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x15 + x16 + x17 + x18 + x19 + x21 + x22, factors = 8, data = Factor_new1, scores = "regression", rotation = "varimax")`

FA <-

- `factanal(-x2+x3+x5+x6+x7+x8+x9+x10+x11+x12+x13+x15+x16+x17+x18+x19+x21+x22, factors=7, rotation="varimax", scores="regression", data=Factor_new1)`

> FA

Call:

`factanal(x = -x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x15 + x16 + x17 + x18 + x19 + x21 + x22, factors = 7, data = Factor_new1, scores = "regression", rotation = "varimax")`
B: Squared Multiple Correlation (SMC) on R

Description

The squared multiple correlation of a variable with the remaining variables in a matrix is sometimes used as initial estimates of the communality of a variable.

SMCs are also used when estimating reliability using Guttman's lambda 6 coefficient.

The SMC is just $1 - \frac{1}{\text{diag}(\text{R.inv})}$ where R.inv is the inverse of R.

Usage

```r
smc(R, covar=FALSE)
```

Arguments

- **R**: A correlation matrix or a dataframe. In the latter case, correlations are found.
- **covar**: If covar = TRUE and R is either a covariance matrix or data frame, then return the smc * variance for each item

Value

- a vector of squared multiple correlations. Or, if covar=TRUE, a vector of squared multiple correlations * the item variances
- If the matrix is not invertible, then a vector of 1s is returned

Author(s)
See Also

mat.regress, factor.pa

Examples:

R <- make.hierarchical()

round(smc(R),2)
References:

C R Kothari, Dr. Research Methodologies: Methods and techniques. 2ed. New Delhi: New Age International (P) Ltd; 2005

Bridget Someth & Cathy Lewin, Research Methods In the Social Science. SAGE Publication Ltd; 2005


Alison Currie et al, Is the child Health/Family Income Gradient Universal? Evidence from England, IZA DP No.1328, October 2004


Luiz Moutinho, University of Glasgow and Graeme Hutcheson, University of Manchester: Statistical Modelling for Business and Management, J.E. Cairnes School of Business and Economics, National University of Ireland Galway: www.Research-Training.net/Galway: June 28-30, 2010
