IJBPAS, November, 2015, 4(11): 6397-6402

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

SUMMARIZING AND CLASSIFICATION OF DATA CONSIDERING COVERED VARIABLES USING FACTORS ANALYSIS TO INVESTIGATE THE GENERAL HEALTH OF PREGNANT WOMEN

ZEYNAB AVAZZADEH

Department of Epidemiology and Biostatistics, School of Public Health, Zahedan University of Medical Sciences, Zahedan, Iran

E Mail: zeynab.evazzadeh@gmail.com

ABSTRACT

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved, uncorrelated variables called factors. In this study, the questionnaire data based on a field of midwifery is used to investigate the general health of pregnant woman (The data has been obtained by "School of Nursing and Midwifery of Mashhad" from "Imam Reza" Hospital).

Keywords: Factor analysis, Midwifery, Eigenvalue, Orthogonal matrix

INTRODUCTION

Factor analysis is used to uncover the latent structure (dimensions) of a set of variables. It reduces attribute space from a larger number of variables to a smaller number of factors. The sample size of research was 55. Those were either pregnant or have given birth less than 6 months. We focus to reduce the data (dimension of data) and summarize data set to fulfill the minimum dimension of data by using the factor analysis. Also, we applied SAS software.

2 Basic of idea

Suppose we have a set of p observable random variables $x_1, x_2, \ldots, x_p$, with means $\mu_1, \mu_2, \ldots, \mu_p$. suppose for some unknown constants l_{ij} and k unobserved random variables F_j, where $i \in 1,2,\ldots,p$ and $j \in 1,2,\ldots,k$ where k<p, we have

$$x_i - \mu_i = l_{i1}F_1 + \cdots + l_{ik}F_k + \varepsilon_i$$

Here, the ε_i are independently distributed error terms with zero mean and finite variance, which may not be the same for all i . Let Var(ε_i)= ψ_i, so that we have

$$cov(\varepsilon) = Diag(\psi_1, \psi_2, \dots, \psi_p) = \psi \quad and \quad E(\varepsilon) = 0,$$

In matrix terms, we have

$$\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$$

If we have n observations, then we will have the dimensions $x_{p\times n}$, $L_{p\times k}$and $F_{k\times n}$. Each column of X and F denote values for one particular observation, and matrix L dose not vary across observations.

Also we will impose the following assumptions on F.

1. $F$and$\underline{\varepsilon}$ are independent.
2. E(F)=0
3. $Cov(F) = 0$

Any solution of the above set equations following the constrains for F is defined as the factors, and L as the loading matrix.

Suppose $Cov(X - \underline{\mu})=\sum$. Then note that from the conditions just imposed on F, we have

$$Cov(X - \underline{\mu}) = Cov(L\underline{F} + \underline{\varepsilon})$$

Or

$$\sum = LCov(\underline{F})L^T + Cov(\underline{\varepsilon})$$

Or

$$\sum = LL^T + \psi$$

Definition (Eigenvalues): The eigenvalue for a given factor reflects the variance in all the variables, which is accounted for by that factor.

A factor's eigenvalue may be computed as the sum of its squared factor loadings for all the variables. The ratio of eigenvalues is the ratio of explanatory importance of the factors with respect to the variables. If a factor has a low eigenvalue, then it is contributing little to the explanation of variances in the variables and may be ignored.

Various rotation strategies are proposed in the literature: Varimax, Oblimin, Quartimin, but the most common rotation strategy is the Varimax rotation. The goal of these rotation strategies is to obtain a clear pattern of loadings i.e. the factors are somehow clearly marked by high loadings for some variables and low loadings for other variables. This general pattern is called 'Simple Structure'.

Remark: Equivalently, a matrix Q is orthogonal if its transpose is equal to its inverse:

$$Q^T = Q^{-1}$$

3 Analysis data

We start Principal Components analysis since it is conceptually somewhat simpler than factor analysis. This does not mean that are recommending it; it just for better pedagogical flow. Imagine you are trying to develop a new measure of Midwifery disease. Your measure contains four questions. For each question, the subject is to respond using the following Likert scale:

1. The lowest (worst)
2. At the appropriate level

3. Good
4. The highest (best)

The first category of the diseases includes the following answers: the 5 first questions and the 6 last questions are selected as the first option. For questions 11 and 25, the third option is chosen. The second category containing the following answers: for questions 6 to 10and 50 to 56, the second option is selected, for questions 26 to 49, the 4th option is done.

At the first we see the screen plot (Fig.1). This plot is used to help determine how many factors to keep in the analysis. It is simply a plot of the eigenvalue, against the number of the factor. One looks for break or "elbows" in the curve. In this graph, it is easy to see that factors 1 and 2 are very different 42-44, so two factors should be retained. In our study, we knew from a theoretical perspective that we wanted two factors. The result support our notion.

Table 1: Variance explained by each factor

$Factor1$	$Factor2$
21.8334205	14.2848562

This table (Table.1) shows the variance explained by each factor (just the first two eigenvalues again) and then the communalities of the variables. The communalities represent how much variance in the original variables is explained by the total of all the factors which are kept in the analysis.

The next figure (Fig.2) is a Plot of two factors retained. Note that at the bottom of the Plot is a key to what the letters are. These are basically two clusters of variables.

If the axes could be rotated so that they ran directly through the clusters. It is needed (Orthogonal mode to factors). SAS software would be how to determine just how to the rotation methods.

Definition (orthogonal mode to factors): in factor analysis district, the factors (in our study factor1 and factor2) are independent of each factor.

Rotation is the step in factor analysis that allows you to identify meaningful factor names or descriptions like these. The next table (Table.2) shows the needed instrument for this step.

This table (table.3) maintains the orthogonal (uncorrelated) nature of the factors and tries to get the original variables to load on one of the factors low on the rest.

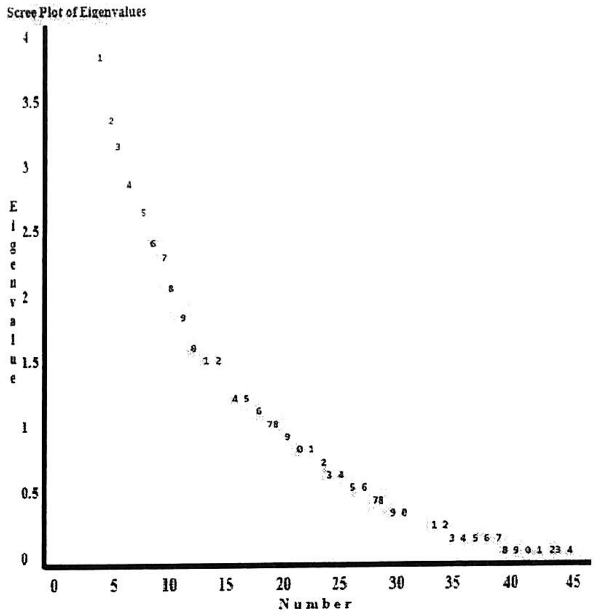

Fig. 1. This plot shows there are many differences between primary factors and finally factors

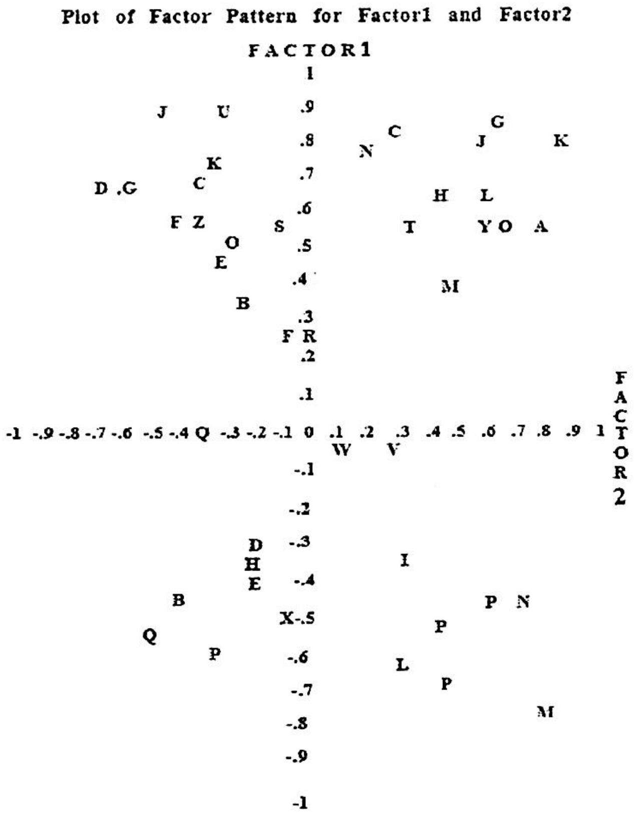

Fig. 2. This plot shows there are many differences between primary factors and finally factors.

quest1=A quest2=B quest3=C quest4=D quest5=E
quest6=F quest7=G quest8=H quest9=I quest10=J
quest11=K quest12=L quest13=M quest14=N quest15=O
quest16=P quest17=Q quest18=R quest19=S quest20=T
quest21=U quest22=V quest23=W quest24=X quest25=Y
quest26=Z quest27=A quest28=B quest29=C quest30=D
quest31=E quest32=F quest33=G quest34=H quest35=I
quest36=J quest37=K quest38=L quest39=M quest40=N
quest41=O quest42=P quest43=Q quest44=R

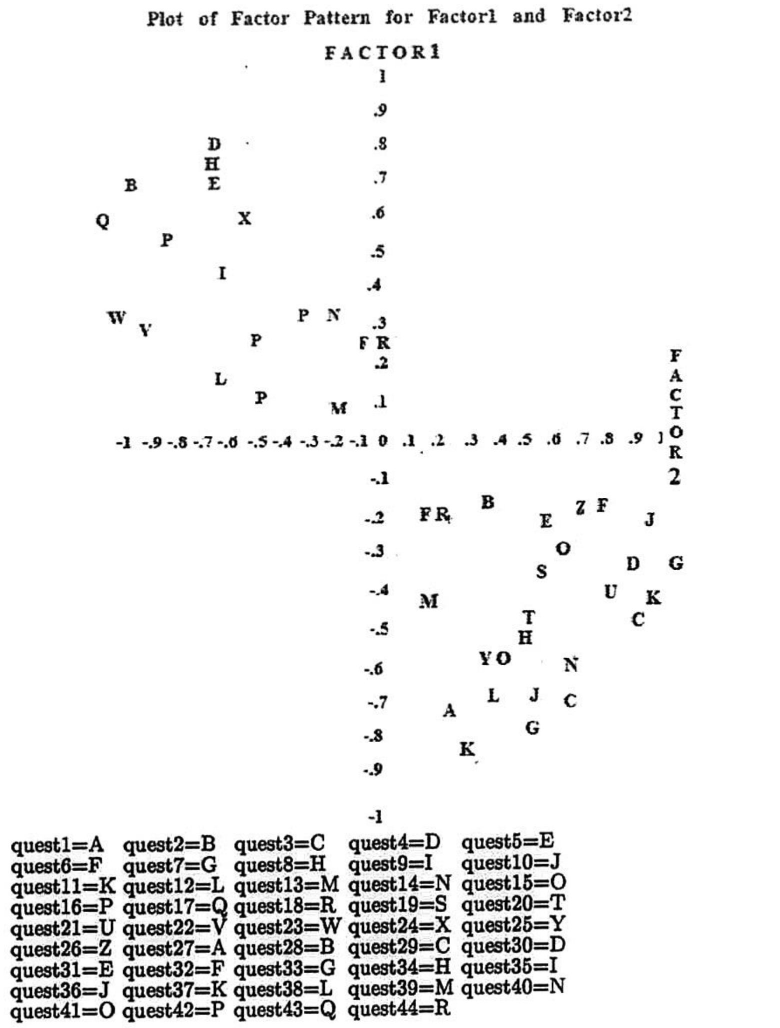

Fig. 3. This plot shows there are many differences between primary factors and finally factors

Table 2: Orthogonal Transformation Matrix

	1	2
1	-0.97699	-0.21329
2	-0.21329	0.97699

Table 3: Variance explained by each factor

Factor 1	*Factor* 2
17.4925103	18.6257664

When this occurs, it is called simple structure. Rotation serves to make the factors more understandable. By seeking so-called "Simple Structure": A pattern loadings where items load most strongly on one factor, and much more weakly on the other factors. Rotations can be orthogonal or oblique (allowing the factors to correlate).

CONCLUSION

In this study here someone cannot be considered to be present in the both of groups. Factor analysis method try to transform the primary variables to the orthogonal factors.

Rotation process is an orthogonal rotation of the factor axes to maximize the variance of the squared loadings of a factor (column) on all the variables (rows) in a factor matrix, which has the effect of differentiating the original variables by extracted factor. Each factor will tend to have either large or small loadings of any particular variable. This solution yields results which make it as easy as possible to identify each variable with a single factor.

The final piece of our analysis (Fig.3) is a plot of the rotated factors. We can see how the new axes really do run right through the clusters.

Notice that the variance explained by each factor has changed. The sum is still the same, but the distribution is more equal now. The communality estimates have not changed since just as much as variance is explained in each original as before rotation.

REFERENCES

1. Cody, R. P. and Smith, J.K. (1997), Applied Statistics and the SAS Programming Language, Prentice-Hall, New Jersey.
2. Johnson, R. A. and Wichern, D. W. (2007) Applied Multivariate Statistical Analysis, 5nd Edition.
3. Cattell, R. B. (2004) the Meaning and Strategic Use of Factor Analysis. Handbook of Multivariate Experimental Psychology, 2nd Edition.
4. Wang, F. (2009), Factor Analysis and Principal – Components Analysis, International Encyclopedia of Human Geography, 1-7.
5. Wu, J. and Bentler, P.M. (2011) Application of H-likelihood to factor analysis models with binary response data, Journal of Multivariate Analysis, vol.106 pp: 72–79.