

# VARIABLE SELECTION AND TEST OF COMMUNALITY IN EXPLORATORY FACTOR ANALYSIS

A. HARADA AND Y. KANO

OSAKA UNIVERSITY, JAPAN

harada@hus.osaka-u.ac.jp, kano@hus.osaka-u.ac.jp

Variable selection in exploratory factor analysis is an important process in the study of human sciences. To identify a variable that may be inconsistent with the factor analysis model considered, there are some criteria such as magnitude of communalities and factor loadings, and model fitting. See e.g., Gorsuch (1988). Deletion of variables with low communalities does not work in some cases, because it is difficult to see whether the insubstantial communality and factor loadings of a certain variables  $X_i$  are caused by the inconsistency of the variable  $X_i$  itself or by the inconsistency of variables other than  $X_i$  which make the communality and factor loadings of  $X_i$  estimated as unduly low values. It is sensible to choose variables with model fit. It is, however, difficult to know which variable should be deleted or appended from output of standard statistical programs.

To help practitioners cope with these problems, Kano and Harada (2000) programmed the SEFA (Stepwise Variable Selection in Exploratory Factor Analysis) system.<sup>1</sup> It is a web-based system and works through the web technology of "Common Gateway Interface (CGI)." The requirement for user's computer is only to connect to the Internet and to prepare a web browser such as 'Netscape Communicator' or 'Internet Explorer.'

The SEFA prints communalities and predicted chi-square goodness-of-fit statistics of each of the models obtained by deletion or addition of one variable in the "Communality and Fit Statistics" section. The SEFA employs model fit as a criterion of variable selection, and one can choose variables that form a well-fitted factor analysis model. Once a well-fitted factor analysis model is obtained, it should not take place that due to an inconsistent variable, the communality and/or factor loadings of some other variables are estimated to be so small. The insubstantial communality surely indicates that the effect of the common factors on the variables is not important.

The program SEFA has a drawback in that it can not identify variables which are fitted well but whose variations are little explained by common factors, i.e., low communalities. In an extreme case, SEFA can not reject a variable that is uncorrelated with the other variables.<sup>2</sup>

To overcome the problem of SEFA, we provide the function to perform test of zero communality for each variable. The theoretical basis was given in Ichikawa (1992), where the asymptotic distribution of the communality estimator was explicitly obtained. The test is formed as a Wald type statistic.<sup>3</sup>

In the talk, we will demonstrate how the enhanced SEFA works for variable selection with low-communality variables.

**Key words:** variable selection, exploratory factor analysis, communality, world wide web.

## References

- Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 231-258.) New York and London: Plenum Press.
- Ichikawa, M. (1992). Asymptotic distributions of the estimators of communalities in factor analysis. *Psychometrika*, 57(3), 399-404.
- Kano, Y. & Harada, A. (2000). Stepwise variable selection in factor analysis. *Psychometrika*, 65(1), 7-22.

---

<sup>1</sup>The analysis with the SEFA system requires correlation matrix, sample size, the number of factors, and the number of variables. One can choose the method of factor extraction and the method of factor rotation, and also can specify statistics to be displayed and coloring. These data are inputted or selected in the form on the web page below:

<http://koko15.hus.osaka-u.ac.jp/harada/sefa2001/stepwise/>

The SEFA system was originally developed in 1997 and was revised in 2000 in which the SEFA was modified to allow for improper solutions.

<sup>2</sup>This fact was pointed out by Mr. Shojima (Waseda University) at the Annual Meeting of the Japan Psychological Association in 2000. The authors thank him for his comment.

<sup>3</sup>The distribution of the test statistic is approximated by the normal distribution. It may cause a problem because the null hypothesis is located at the boundary of the parameter space. This problem will be discussed elsewhere.