A FACTOR ANALYTIC APPROACH TO DISCARDING VARIABLES

A. C. Rencher, Brigham Young University

SUMMARY

Often, when a large number of variables are observed, little information is lost if some of the variables are discarded. A method based on factor analysis is proposed for selecting the variables to be rejected. An example is given wherein the method is used to reduce the number of questions on a questionnaire.

1. INTRODUCTION

In some multivariate situations it is desirable or even necessary to reduce the number of variables and work with a "good" subset. A method such as principal component analysis could be used to obtain a set of new variables defined as linear combinations of the original variables. Often a relatively small set of such new variables will preserve most of the information. However, in some cases the new variables are not useful. For example, suppose it is required to reduce the size of a questionnaire or test. It is usually not practicable to use new questions defined as linear combinations of the original ones. Even if such new questions could be identified and phrased meaningfully a problem arises in their use. If the subjects are asked to respond to the new questions it will be found that their answers are correlated. Much information is thereby lost since the new variables would be independent if obtained as orthogonal functions of the original variables.

In most common multivariate situations, the intended analysis is of greater interest than a prior reduction in the number of variables. However, in many cases the results will be only slightly affected if certain variables are discarded before performing the analysis. The ensuing savings in the number of required measurements would be especially helpful in situations which recur routinely.

2. ALTERNATIVE SELECTION METHODS

Several methods have been suggested for selection of variables. The following three approaches are discussed extensively by Jolliffe (1972, 1973).

(1) A stepwise multiple correlation method which discards successively the variable having the largest multiple correlation with the remaining variables. An empirically determined stopping point is suggested.

(2) Principal component methods which associate a variable with each of the principal components and discard those variables associated with the principal components having the smallest corresponding eigenvalues. An empirically determined cut off point is proposed.

(3) Clustering methods which segregate the variables into groups and select one variable from each of the resulting groups. The clustering stops when an empirically determined termination point is reached.

These three methods together with variations are compared by Jolliffe using both real and artifical data. None of the methods was found to be uniformly superior to the others.

3. A METHOD BASED ON FACTOR ANALYSIS

Intuitively, it seems one should be able to examine the correlation matrix and from the patterns therein discover which variables can be most readily sacrificed as providing the least additional information above that already available in their fellows. The three methods listed in Section 2 are in fact based on extracting information from the correlation matrix. Other simpler but less efficient approaches readily suggest themselves. For example, the sum of squares of each row of the correlation matrix could be examined and the variable with the largest sum of squares deleted. This process could be continued in a stepwise fashion after deletion of the appropriate row and column at each step. It appears however, that this procedure would be inferior to the first method reported in Section 2. It might be considered only if time or computational limitations prescribed its use. A far more satisfactory procedure for examining the correlation matrix is now described.

Consider the usual (orthogonal) factor analysis model for p variables expressed in terms of m factors.

$$\begin{aligned} x_1 &= \lambda_{11}f + \lambda_{12}f_2 + \ldots + \lambda_{1m}f_m + e_1 \\ x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \ldots + \lambda_{2m}f_m + e_2 \qquad (1) \\ &\vdots \\ x_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \ldots + \lambda_{pm}f_m + e_p \end{aligned}$$

In matrix form (1) can be expressed as $\underset{\sim}{x} = \underset{\sim\sim}{\Lambda}\underset{\sim}{f} + \underset{\sim}{e}$. The following will be assumed: $E(\underset{\sim}{x})=\underset{\sim}{0}$, $\mathrm{cov}(\underset{\sim}{x}) = \underset{\sim}{\Sigma}$, $E(\underset{\sim}{f}) = \underset{\sim}{0}$, $\mathrm{cov}(\underset{\sim}{f}) = \underset{\sim}{I}$, $E(\underset{\sim}{e}) = \underset{\sim}{0}$, $\mathrm{cov}(\underset{\sim}{e}) = \underset{\sim}{\Psi} = \mathrm{diag}(\Psi_1,\ldots,\Psi_p)$, and $\underset{\sim}{e}$ and $\underset{\sim}{f}$ are independent. From these assumptions it follows that

$$\underset{\sim}{\Sigma} = \underset{\sim\sim}{\Lambda\Lambda'} + \underset{\sim}{\Psi}. \qquad (2)$$

In practice the correlation matrix is often used in place of $\underset{\sim}{\Sigma}$.

If two variables x_i and x_j have nearly identical loadings on the m factors, they contribute similar information and one of the two may be deleted. The following weighted distance function is suggested as a measure of the degree of closeness or similarity of x_i and x_j:

$$d_{ij} = w_1(\lambda_{i1} - \lambda_{j1})^2 + w_2(\lambda_{i2} - \lambda_{j2})^2 + \ldots + w_m(\lambda_{im} - \lambda_{jm})^2 . \qquad (3)$$

If the λ_{ik} are estimated by factoring $\underset{\sim}{\Sigma}$ (or $\underset{\sim}{\Sigma}-\underset{\sim}{\Psi}$) using eigenvalues and eigenvectors, appropriate weights are provided by the eigenvalues. Thus, w_1 is the largest eigenvalue, w_2 is the next largest, etc. If another method has been used to estimate the loadings or if they have been rotated,

appropriate weights can be found from

$$w_k = \sum_{i=1}^{p} \lambda_{ik}^2 \qquad k = 1, 2, \ldots, m.$$

The following procedure is suggested to determine which variables can best be discarded. Find d_{ij} for each pair x_i and x_j, $i \neq j$. Arrange these in ascending order. Then from the pair of x's with smallest distance function value, retain the one with smaller Ψ value and delete the other. Similarly, one of the two variables with next smallest d_{ij} can be deleted. This process can continue until as many variables as desired are discarded. In each case when choosing between two variables it seems preferable to retain the one with smaller value of Ψ. If two variables are found to be similar and later each is found to be close to a third variable, two of the three may be deleted. This suggests the possible use of a clustering procedure. However, cluster analysis is not recommended due to the problem to be discussed next which can best be handled in the pairwise framework.

Negative correlations have an effect on the location of points in the space of factor loadings. If the response on a question is given on a multipoint ordered scale between poles, a reversal of the poles will change the sign of the correlation of the given question with all others. From (2) it is clear that

$$\sigma_{ij} = \sum_k \lambda_{ik}\lambda_{jk}, \; i \neq j.$$

A change in sign produces

$$-\sigma_{ij} = \sum_k (-\lambda_{ik})\lambda_{jk} = \sum_k \lambda_{ik}(-\lambda_{jk}).$$

Thus, either $(\lambda_{i1},\lambda_{i2},\ldots,\lambda_{im})$ or $(\lambda_{j1},\lambda_{j2},\ldots,\lambda_{jm})$ has all signs reversed and appears on the opposite side of the origin from its original position. If two points are close together in the factor loading space, a reversal in sign of the correlation between the two variables they represent would place the two points very far apart. The distance function d_{ij} would then fail to detect the close similarity between the two variables. Allowance can be made for this possibility when making each comparison by considering $\Sigma\, w_k(\lambda_{ik} + \lambda_{jk})^2$ as well as $\Sigma\, w_k(\lambda_{ik}-\lambda_{jk})^2$ in each case and choosing the smaller as the distance between the two variables in question. Thus (3) becomes

$$d_{ij} = \min \{\Sigma w_k(\lambda_{ik}-\lambda_{jk})^2, \; \Sigma w_k(\lambda_{ik}+\lambda_{jk})^2\}. \quad (4)$$

If the number of variables to be discarded has not been predetermined, some visual assistance in arriving at an appropriate value can be obtained by ordering the distances and plotting them. The plot can then be examined for a turning point where a noticeably steeper ascent begins.

4. EXAMPLE

A teacher evaluation survey was administered periodically to determine students' rating of their university teachers. The major part of the questionnaire consisted of 22 questions about various aspects of teacher effectiveness. When a revision was contemplated, information was sought as to whether some of the questions could be deleted with little sacrifice in information. A shorter questionnaire would meet with less student resistance.

A factor analysis was performed using a correlation matrix obtained from past survey results. It was found that 17 factors were required to explain 95% of the variation. The distance function (4) was calculated for each of the 231 pairs of variables. The smallest 20 of these are given in Table 1. The remaining distances ranged up to a high of 2.875.

TABLE 1
Pairs of questions with smallest distance functions

Question Pair	Distance	Question Pair	Distance
19,20	.023	14,19	.657
21,22	.041	2,19	.797
8,20	.151	2,20	.877
8,19	.170	5,19	1.016
1,16	.172	8,14	1.072
2, 7	.174	7,19	1.082
8,10	.352	5,20	1.112
10,19	.453	2,14	1.113
10,20	.473	2, 8	1.132
14,20	.603	13,19	1.163

Note that the pair of variables that are closest, 19 and 20, are also both very close to 8. Thus, only one of these three variables need be retained.

As an aid to determining how many questions might appropriately be deleted, the first 15 distances are plotted in Figure 1. There is a definite change in pattern from the sixth to seventh value. It seems clear that a variable can be deleted from each of the first two pairs with almost no loss of information. The next four distances also appear small enough to warrant deletion of one variable from each pair if desired. Beyond that point, however, further deletion may not be justified.

(FIGURE 1 HERE)

5. DISCUSSION

The methods discussed by Jolliffe (see Section 2) appear to be well suited to situations where a rather small subset of the original variables is adequate. The present method is recommended where relatively few (up to one-third, say) of the variables are to be discarded. This situation can be readily identified when the factor analysis indicates that a rather large number of factors are required to "explain" the data.

The method of this paper has the added advantage of simplicity of operation. It is easily programmed and can be included as an option in a standard factor analysis routine. The resulting ordered distance values can be examined meaningfully by someone with little statistical expertise.

The number of factors to be used is somewhat arbitrary. Good results will be obtained by retaining at least enough to "account for" about 95% of the variance in the system.

REFERENCES

JOLLIFFE, I. T. (1972). Discarding variables in a principal component analysis. I: Artificial data. Appl. Statist., 21, 160-173.

JOLLIFFE, I. T. (1973). Discarding variables in a principal component analysis. II: Real data. Appl. Statist., 22, 21-31.

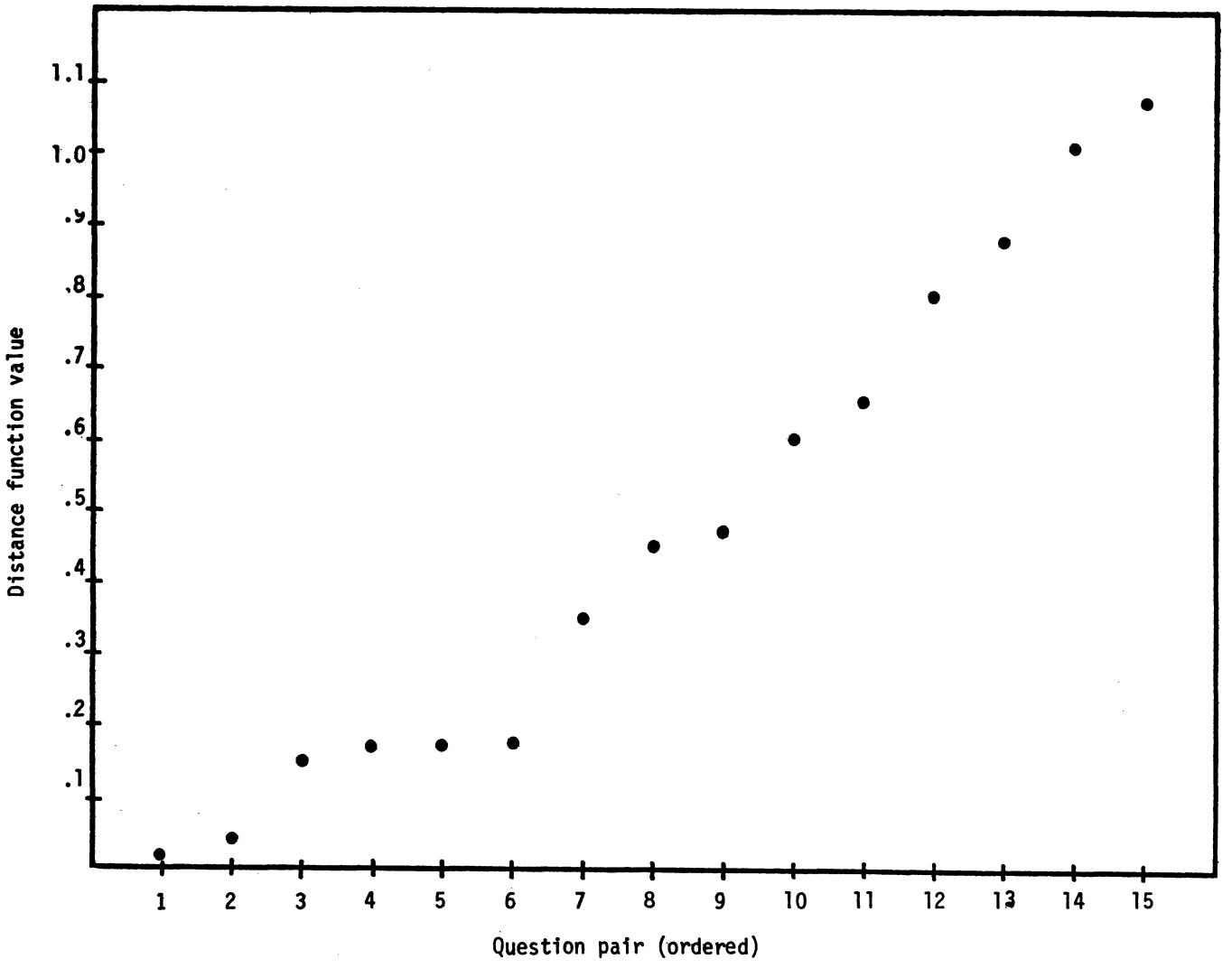

Figure 1. The 15 smallest distance values.