This paper describes the authors' FORTRAN algorithm FACAIC for choosing the number of factors for an orthogonal factor model using Akaike's Information Criterion. FACAIC utilizes the IMSL subroutine OFCOMM.

Key words: factor analysis algorithm, choosing the number of factors, AIC and CAIC.

Purpose and Description

Purpose

Factor analysis is a very useful and important multivariate technique which is used to find a way of condensing or summarizing the information contained in a number of original variables into a smaller set of composite dimensions, factors, or hypothetical constructs with minimum loss of information.

In the statistical literature, an exact satisfactory quantitative technique for deciding how many factors to extract is largely unresolved, and does not seem to be available, except some of the ad hoc stopping rules which are currently being utilized. The classical theory of hypothesis testing is not appropriate in the context of factor model problems, since the problem is not of testing a particular hypothesis, but is rather a multiple decision problem. The usual likelihood ratio test is not always valid, and the significance level for the test criterion is not adjusted in the sequential testing process.

Therefore, the algorithm presented in this paper uses Akaike's (1974) information criterion (AIC) and the asymptotically consistent Akaike's information criterion (CAIC) to choose the number of factors. These being information-theoretic identification procedures, the researcher is forced to be more neutral in the choice of the model and the naming of the common factors.

Description

The factor analysis model assumes a structure of the form:

\[ X = \Lambda f + \mu + \varepsilon \]  

(1)

where

\[ \Lambda = \text{the factor loading matrix, that is, a matrix of unknown coefficients } \lambda_{ij}, \]

\[ f = \text{a random vector of common factors,} \]

The authors dedicate this algorithm to Professor Hirotugu Akaike in appreciation of his pioneering work on AIC which was originally intended for the factor analysis and other statistical model identification problems. Requests for reprints should be sent to Hamparsum Bozdogan, Department of Mathematics, University of Virginia, Charlottesville, VA 22903.
\[ \mu = \text{mean vector of } p \text{ variables,} \]
\[ \varepsilon = \text{a random vector of specific or unique factors, or the error term,} \]

and with

\[ f \text{ and } \varepsilon \text{ independent} \]
\[ E(f) = 0, \]
\[ \text{Cov}(f) = I, \]
\[ E(\varepsilon) = 0, \]
\[ \text{Cov}(\varepsilon) = \psi, \text{ where } \psi \text{ is a diagonal matrix.} \]

The algorithm is presented as a FORTRAN 77 subroutine. The user is assumed to have computed the correlation matrix \( R \) from the data matrix \( X \). The user inputs to the subroutine FACAIC are:

- \( N \) : INTEGER
  NUMBER OF CASES
- \( P \) : INTEGER
  NUMBER OF VARIABLES
- \( DOAIC \) : LOGICAL
  .TRUE. FOR AIC
  .FALSE. FOR CAIC
- \( CORR \) : DOUBLE PRECISION VECTOR OF DIMENSION \( (P^*(P + 1)/2) \)
  CORRELATION MATRIX IN SYMMETRIC STORAGE
- \( AUTO \) : LOGICAL
  .TRUE. FOR AUTOMATIC SEARCH FOR BEST MODEL
  .FALSE. FOR MODEL PREDETERMINED BY USER
- \( SETNF \) : LOGICAL
  NUMBER OF FACTORS PREDETERMINED BY USER
  WHEN AUTO = .FALSE.

The outputs of the subroutine FACAIC are:

- \( BESTNF \) : INTEGER
  BEST NUMBER OF FACTORS
- \( ALPHA \) : DOUBLE PRECISION
  \( P \)-VALUE FOR GOODNESS-OF-FIT
- \( NPARMS \) : DOUBLE PRECISION
  NUMBER OF PARAMETERS
- \( PENALT \) : DOUBLE PRECISION
  PENALTY COMPONENT FOR AIC OR CAIC USING \( BESTNF \) FACTORS
- \( AIC \) : DOUBLE PRECISION
  AKAIKE'S INFORMATION CRITERION FOR MODEL SELECTION