ON THE STATISTICAL TREATMENT OF RESIDUALS IN FACTOR ANALYSIS*

K. G. Jöreskog

UNIVERSITY OF UPPSALA, SWEDEN

A method for estimation in factor analysis is presented. The method is based on the assumption that the residual (specific and error) variances are proportional to the reciprocal values of the diagonal elements of the inverted covariance (correlation) matrix. The estimation is performed by a modification of Whittle's least squares technique. The method is independent of the unit of scoring in the tests. Applications are given in the form of nine reanalyses of data of various kinds found in earlier literature.

In factor analysis when $p$ tests are given to $n$ individuals we get a matrix of test scores

$$S = [s_{iv}], \quad i = 1, 2, \ldots, p; \quad v = 1, 2, \ldots, n.$$ 

The fundamental postulate of factor analysis ([35], p. 63) is expressed in the linear equation

$$s_{iv} = \alpha_{i1}\beta_{1v} + \alpha_{i2}\beta_{2v} + \cdots + \alpha_{ik}\beta_{kv} + \epsilon_{iv},$$

or in matrix form

$$S = \alpha\beta + \epsilon,$$

where

$$\alpha = [\alpha_{iv}], \quad \beta = [\beta_{iv}], \quad \epsilon = [\epsilon_{iv}].$$

We shall regard both $\alpha_{iv}$ and $\beta_{iv}$ as nonrandom quantities. The matrix $\epsilon$ is a set of random variables, about which we shall assume only that

$$(3a) \quad E(\epsilon) = 0,$$

$$(3b) \quad E(\epsilon\epsilon') = n \Delta,$$

where $\Delta$ is a positive-definite diagonal matrix. The diagonal elements of $\Delta$ will be called residual variances. (Thurstone [35] uses the term uniqueness after standardization of the test scores.)

The most familiar interpretation of this model is in terms of mental tests. In the battery of the $p$ tests there are $k$ common factors. The $\alpha_{iv}$ are the factor loadings of these factors, the assumption being that the loadings

*The writer wishes to thank Prof. H. Wold, Dr. E. Lyttkens, and Dr. P. Whittle for valuable comments and suggestions.
are invariant under changes of the population of individuals. The $\beta_{it}$ are
the factor values of the individuals, i.e., their latent scale values on the
different factor scales. We shall assume, more for convenience than necessity,
that these mutual scales have been fixed so that

\begin{equation}
\frac{1}{n} \sum_{i=1}^{n} \beta_{it} = 0, \quad t = 1, 2, \ldots, k,
\end{equation}

\begin{equation}
\frac{1}{n} \sum_{i=1}^{n} \beta_{is} \beta_{tt} = \begin{cases} 0 & \text{if } s \neq t, \\ 1 & \text{if } s = t. \end{cases}
\end{equation}

The $\epsilon_{it}$, which will be called residuals, can be interpreted as the sum of an
error and a specific factor.

The observed moment matrix $C$ will be, using (2) and (4b),

\begin{equation}
C = \frac{1}{n} SS' = \alpha \alpha' + \alpha \beta \frac{e'}{n} + \frac{e}{n} \beta' \alpha' + \frac{ee'}{n}.
\end{equation}

Taking expectations on both sides of (5) and using (3a) and (3b) we get

\begin{equation}
\Sigma = \alpha \alpha' + \Delta.
\end{equation}

In (6) we have denoted $E(C)$ by $\Sigma$.

A major statistical problem of factor analysis is to estimate $\alpha$ from the
observed test scores $S$ or the observed covariances $C$. For certain applications
a knowledge of the factor values $\beta$ is also useful. We can distinguish between
at least three different kinds of approaches to the statistical problems in
factor analysis.

Among psychologists the currently most preferred approach is to esti-
mate $\alpha$ according to some communality method. The diagonal elements
of $\Sigma - \Delta$ are called communalities. (Thurstone [35] uses that term after
standardization of the test scores.) One wants to choose the communalities
so that $\Sigma - \Delta$ is positive semidefinite and of smallest possible rank. The
estimation procedure is in two steps. First we choose some diagonal matrix
$D$ which will make $C - D$ positive semidefinite and approximately of mini-
mum rank. Then $C - D$ is factored to give

\begin{equation}
C - D = AA'.
\end{equation}

$A$ is then taken as an estimate of $\alpha$. This method is nonstatistical in nature
since it ignores the statistical variation. From the statistical point of view
we should choose $D$ so that $C - D$ is positive semidefinite and of minimum
rank within possible statistical variation, and then we should choose some
$A$ so that $C - D - AA'$ is a zero matrix within possible statistical variation.
As far as the writer knows there is no statistical basis for the first of these
two steps. Anderson [1] has reviewed three methods for determining $A$ in
the second step, when the first step is completed. These methods are the