TO WHAT EXTENT CAN COMMUNALITIES REDUCE RANK?

LOUIS GUTTMAN
THE ISRAEL INSTITUTE OF APPLIED SOCIAL RESEARCH AND
THE HEBREW UNIVERSITY IN JERUSALEM

The question is raised as to whether the null hypothesis concerning the number of common factors underlying a given set of correlations should be that this number is small. Psychological and algebraic evidence indicate that a more appropriate null hypothesis is that the number is relatively large, and that smallness should be but an alternative hypothesis. The question is also raised as to why approximation procedures should be aimed primarily at the observed correlation matrix $R$ and not at, say, $R^{-1}$. What may be best for $R$ may be worst for $R^{-1}$, and conversely, yet $R^{-1}$ is directly involved in problems of multiple and partial regressions. It is shown that a widely accepted inequality for the possible rank to which $R$ can be reduced, when modified by communalities, is indeed false.

When Charles Spearman hypothesized, some fifty years ago, that correlations among certain mental test scores could be accounted for by but a single common factor, this was greeted by many of his colleagues with substantial scepticism. In current terminology, initiated by L. L. Thurstone, they thought it implausible that communalities could be found that would reduce the given type of observed correlation matrix to rank one. Thurstone later hypothesized that relatively small rank could be attained for correlation matrices of mental test data by use of communalities. This hypothesis, too, has encountered a measure of disbelief in various quarters.

Motivation for seeking small rank stems from the desire to reproduce the observed correlations among $n$ variables by using scores on a smaller number, say $m$, of common factors. A necessary and sufficient condition that there exist $m$ common factors to do the reproducing trick is that there exist communalities that reduce the rank of the observed correlation matrix to $m$, but leave the matrix Gramian [8]. In this sense, rank and number of common factors are equivalent. It is algebraically convenient to deal with rank in studying the possible number of common factors for a given set of data.

Empirical attempts to estimate minimal rank in given cases have hitherto not been clear cut for lack of rigorous theory and computing routines for fallible data. Among the better efforts in this direction are the works of Lawley [13] and Rao [16]. But these do not presume to do more than give a lower bound to the minimal rank; they do not provide any upper bounds.

*This research was facilitated by a noncommitted grant-in-aid to the writer from the Ford Foundation.
Regardless, if all results published to date be taken at their face value, they together constitute abundant evidence against the Thurstone hypothesis. Hundreds of different common factors for mental abilities have been "identified" by Thurstonian computing routines, and the number is still growing, with no upper limit in sight (cf. [2] and the discussions in [12] and [19]).

The association of the notion of communality coefficients with that of minimal rank seems to be an historical accident, as it is not logically necessary [4, 8, 9, 10]. Identifying the concept of small rank with that of scientific parsimony also seems fortuitous, as this too has no logical compulsion; other kinds of parsimony are possible [6, 12]. In view of the evidence favoring large rank—even when communalities are used—for the entire domain of mental abilities, it is fortunate for the communality concept that it is useful and meaningful regardless. For example, a unique definition of communalities is sometimes possible in terms of image analysis, without considering rank at all [4, 9, 10]. There are other possibilities for other cases, especially where facet design is used for the tests [12]. Communalities can lead to parsimonious descriptions of data without reference to rank.

Individual research projects, however, typically study but a small sector of the domain of mental abilities and not the entire domain. Is it still not true that "in general" communalities will meaningfully reduce rank in the subdomains? Also, isn't it an algebraic fact that, psychological meaning aside, communalities for a given correlation matrix of order n can always be found to reduce the rank to m, where the inequality is satisfied,

\[ m \leq \frac{1}{2}(2n + 1 - \sqrt{8n + 1}) \]

as was shown by several authors [1, 14, 18]?

The purpose of the present paper is to explore these last questions. Despite the rather widespread belief in inequality (1), it can be proved false. Its authors make a valiant attempt to establish an algebraic upper bound to minimal rank, but introduce an important fallacy. In fact, the best possible universal upper bound to minimal m is \( n - 1 \), as shall be shown.

Further algebraic evidence will be submitted against small rank being the general algebraic case. For fallible data, this implies that large rank should be the null hypothesis, not to be rejected unless there is strong empirical evidence to the contrary. Most current computing routines have the opposite orientation; their justification accordingly may require reexamination. Because of the confusion as to the nature of the rank hypothesis to be tested by fallible data, we open our discussion of communalities with this problem of orientation. Psychological considerations will be stressed no less than purely algebraic ones.