Recommended citation:
McDonald, R. P. (1967). *Nonlinear Factor Analysis* (Psychometric Monograph No. 15). Richmond, VA: Psychometric Corporation. Retrieved from http://www.psychometrika.org/journal/online/MN15.pdf

PSYCHOMETRIC MONOGRAPH NUMBER 15

NONLINEAR FACTOR ANALYSIS

The William Byrd Press, Inc.

Richmond, Va.

NONLINEAR FACTOR ANALYSIS

Roderick P. McDonald

UNIVERSITY OF NEW ENGLAND
N. S. W., AUSTRALIA

ACKNOWLEDGMENTS

The present monograph, except for minor revisions, was presented as a doctoral dissertation at the University of New England in 1963. I wish to thank my supervisor, Dr. J. A. Keats of the University of Queensland, for his encouragement, advice, and criticism in the course of this work.

I am particularly indebted to Dr. E. J. Burr, of the Department of Mathematics, University of New England. His advice and criticism at all stages in the work has helped me to avoid a number of lapses from formality and rigor in the mathematics to be presented. Any errors that remain are, of course, my responsibility.

The final draft of the manuscript was prepared while I was a Visiting Research Fellow at Educational Testing Service.

NONLINEAR FACTOR ANALYSIS

RODERICK P. MCDONALD

UNIVERSITY OF NEW ENGLAND, N. S. W., AUSTRALIA

NOTE ON FORMAT AND NOTATION

Sections, equations, tables and figures, are numbered on Peano's decimal system. Thus, (4.8.2) means the second equation of section 8 of Chapter 4, and similarly for tables and figures. The sections of Chapter 5 correspond to some of the sections of Chapter 4, and are given section numbers identical with those of the sections in Chapter 4 on which they are based.

An attempt has been made to conform to the preferred conventions in mathematical notation. In general, capital Roman letters (in italics) denote matrices, lower case Roman letters (in italics) without subscripts denote vectors, and Greek letters denote scalars. The few exceptions should be clear from the context. (For example, a script R ($\mathfrak{R}$) is used for a covariance matrix, and the corresponding correlation matrix is the ordinary R.)

A script E ($\mathcal{E}$) is employed for the expectation operator.

A prime after a matrix or vector denotes transposition.

Significance levels: $*P < .05$, $**P < .01$, $***P < .001$.

Glossary of major symbols (with equation of first appearance).

(2.1.1) $y \equiv [y_1 \cdots y_n]$, observed random vector.

(2.1.1) $x \equiv [x_1 \cdots x_r]$, vector, standardized "factor scores."

(2.1.1) $e \equiv [e_1 \cdots e_n]$, vector, unique deviations.

(2.1.1) A, an $r \times n$ matrix, "common factor loadings."

(2.1.1) U, an $n \times n$ diagonal matrix, "unique factor loadings."

(2.1.3) $\mathfrak{R}$, an $n \times n$ matrix, observed covariances.

(2.1.4) M, an $r \times n$ matrix, normalized latent vectors.

(2.1.4) C, an $r \times r$ diagonal matrix, latent roots.

(2.1.6) $F \equiv C^{1/2}M$, an $r \times n$ matrix, principal component loadings.

(2.1.7) L, an $r \times r$ orthogonal matrix, rotation.

(2.1.14) K, an $n \times n$ diagonal matrix, variances of $y_1, \cdots, y_n$.

(2.1.14) z, vector obtained by standardizing y.

(2.1.15) R, an $n \times n$ matrix, observed correlations.

(2.3.1) $f(y)$, joint density function of $y_1, \cdots, y_n$.

(2.3.1) ψ, unspecified latent "characterization."

(2.3.1) $g(\psi)$, density function of ψ.

(2.3.1) $h(y \mid \psi)$, conditional density function of y given ψ.

(2.3.2) $h_i(y_i \mid \psi)$, conditional density function of any y_i given ψ.

(2.3.7) $p_k \mid \psi$, conditional probability that a dichotomous $y_k = 1$, given ψ.

(2.3.14) $$N(\cdot) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty} \exp\left(-\frac{x^2}{2}\right) dx,$$

the normal ogive function.

(2.3.17) $f_i(\Psi) \equiv \mathcal{E}(y_i \mid \Psi)$, a prescribed function.

(2.3.21) p_i, probability of a "positive" response.

(2.3.21) q_i, probability of a "negative" response.

(2.5.7) $h_p(x)$, $p = 1, \cdots, n$, orthogonal polynomials in x.

(3.1.5) $q \equiv [Q_1(x_1, \cdots, x_t) \cdots Q_r(x_1, \cdots, x_t)]$, a vector of r linearly independent functions in t statistically independent variates x_m.

(3.1.8) T, an $r \times r$ nonsingular matrix.

(3.1.9) $h \equiv qT \equiv [h_1(x_1, \cdots, x_t) \cdots h_r(x_1, \cdots, x_t)]$, a vector of r orthonormal functions in the x_m.

(3.1.12) B, an $r \times n$ matrix, "factor loadings."

(3.1.20) $v \equiv [v_1 \cdots v_r] = zF'C^{-1}$, a vector of "component variates."

(3.1.20) $w \equiv [w_1 \cdots w_r] = hL$, a vector of "true parts" of the "component variates."

(3.1.20) $d \equiv [d_1 \cdots d_r] = eUF'C^{-1}$, a vector of "disturbances."

(3.1.25) S, an $r \times r$ matrix, covariance matrix of d.

(3.2.1) $\mu_p \equiv \mathcal{E}(x^p)$, moments of a single factor x.

(3.2.3) Φ_p, $p = 2, 3, \cdots r$, functions in the moments of w, and initially unknown parameters comprising elements of a rotation-matrix and the moments of the latent distribution.

(3.7.1) c_{ip}, Fourier coefficients defined by (3.7.2).

(4.4.1) $$n(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right),$$

the normal density function.

(4.8.3) Y, an $n \times m$ matrix of observations on n "individuals" under m "levels of a quantitative treatment."

(4.8.3) $F(x)$, an $s \times m$ matrix.

(4.8.3) E, an $n \times m$ matrix of "errors" in Y.

(4.8.7) $G(x)$, an $s \times m$ matrix, obtained by orthonormalizing $F(x)$.

(4.8.19) K, an $n \times (\ell + 1)$ matrix of coefficients.

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

The object of the following is to present an account of a general system for nonlinear factor analysis, to illustrate its usefulness with the aid of a number of empirical applications and some constructed numerical examples, and to show how the concepts employed in the system lead to a degree of theoretical unification of the field of psychometric models for the dimensional analysis of data.

In a broad historical perspective, until the early 1940's there was only one model for the dimensional analysis of multivariate data. This was the linear factor analysis model stemming from the work of Pearson and developed further by Burt, Spearman and Thurstone, to mention only the most outstanding names. About the only alternative approach to problems of dimensionality, up to this time, concerned methods for constructing relatively homogeneous tests. With few exceptions, such methods were so lacking in rigor that they are best regarded as providing recipes rather than models.

A besetting problem of this earlier period concerned the application of factor-analytic concepts and methods to "category" or "qualitative" data. It was felt that factor-analytic concepts should provide a rational basis for the construction of unidimensional tests. However, this would involve operating on observations obtained from multicategory or dichotomous items, whereas the basic factor equations, it was thought, had been developed in terms of continuously measurable observations. The issue of the applicability of factor analysis to items became sharpened by the recognition that there was a choice of coefficients of association between items, as against the *apparently* God-given nature of the product-moment correlation coefficient for "quantitative" variables. Within the traditional framework, this led to a discussion of the relative merits of the alternative coefficients for items, in terms of the possibility that factoring these would yield artifacts in the form of "difficulty factors," or lead to embarrassments such as negative latent roots and nonreal factor loadings.

More importantly, the problem of qualitative observations led to a search for alternative methods of dimensional analysis, outside the factor-analytic tradition. Major developments were the refinement by Ferguson and Guttman of existing notions of consistent answer patterns into the concept of the perfect scale, and the development by Lazarsfeld, in a somewhat strange new language, of the concepts of latent structure analysis. Latent structure analysis has provided a fairly general framework which

can both subsume existing models as particular cases and proliferate further models according to taste. Further, by a generalization on his work on the principal components of the perfect scale, Guttman has more recently developed some new models for quantitative observations (the radex theory), which he contrasts with the Spearman-Thurstone model.

The present situation, then, is in marked contrast to the period up to the early 1940's. Whereas in the early period the research worker had one model for the intrinsic analysis of his test data, applied with or without misgivings to qualitative or quantitative observations, there is at the present time a whole range of models to choose from. In fact we seem to have an *embarras de richesse*. To take a particular example, consider a research worker confronted with a matrix of observations consisting of dichotomous responses from a large number of subjects to a questionnaire of, say, ten or twenty items. He may consider *at least* the following modes of analysis, each requiring a different treatment of the data right from the beginning of the analysis: (a) scalogram treatment, (b) the latent distance model, (c) the normal ogive, (d) the latent linear model, (e) the latent hyperplane (an obvious generalization of the latent linear model), (f) the latent class model, (g) Coombs' nonmetric compensatory model, (h) Coombs' conjunctive/disjunctive model, (i) "conventional" methods of item analysis, (j) a "conventional" factor analysis of one of several coefficients of association between items.

The primary object of the work to be reported here is to develop a generalization of Spearman-Thurstone common-factor analysis, to allow of nonlinear relations between observed variables and factors. It turns out that the initial steps in the analytical procedure are identical with those of linear factor analysis up to the point of obtaining an orthogonal factor solution. Thereafter, further analysis can serve to fit to the data, and to justify, the most appropriate linear or nonlinear model. From the theoretical relations between such psychometric models as those mentioned above, and the general model to be treated here, it is possible, at least in principle, to convert the parameters of the nonlinear solution obtained into the parameters of an alternative model. The choice of the "best" alternative model is determined by the nature of the initial nonlinear solution. In this sense, it is a natural consequence of the present approach that it provides a conceptual unification of the field of "dimensional" psychometric models, such as latent structure analysis does. Further than this, it provides a single starting point in a practical job of analysis, with a rational choice among the alternative models at a late stage in the work.

In the chapter immediately following, the background of the present work is sketched in. While part of the object of this chapter is to develop in more detail the historical points that were so briefly indicated above, these are only treated to the extent that they provide useful pointers to the present theory.

In the third chapter, the general theory of nonlinear factor analysis is developed. Following an account of the model in its most general form, a complete theoretical account is given of the "single-factor" polynomial, which may be thought of as a generalization, for quantitative observations, on Lazarsfeld's latent polynomial model for dichotomies, and includes the latter as a special case. The multiple-factor polynomial is then treated completely for cases where no terms are present in the regression functions which involve products of factor scores. Following this, an informal account is given of the type of case where such product terms may occur.

Unfortunately, difficulties in the practical solution of problems by the present methods increase as a positively accelerated function of the approximate rank of the observation matrix. There is reason to be confident that the system can be made practicable and comprehensive with the extensive development of suitable computer programs. The present work should be regarded as a foundation for future developments. On the other hand, the empirical work to be presented will serve to show that the system is already to the point of practical use on at least a limited range of problems. The theory developed in the third chapter is there illustrated by artificially constructed examples only.

The fourth chapter provides a broad, though not exhaustive, sampling of the relations of the system being presented to existing problems and methods.

It may be noted that the theory is just a piece of applied mathematics. Hence it does not stand or fall according to what is the case in behavioral nature. The empirical work presented in the fifth chapter serves to show two things only: that nonlinear relations between "tests" and "factors" can be found in behavioral data, and that the present system can be employed to detect and fit functions to them.

Thus far, empirical examples for nonlinear factor analysis have been sought where theory suggests that they are most likely to be found. Consequently, each such example serves to illustrate one of the specializations of the theory, as developed in the fourth chapter, as well as illustrating the general methods as given in the third chapter. Accordingly, the sections dealing with empirical investigations are labelled in conformity with the sections of the preceding chapter containing the appropriate specializations of the theory.

In the final chapter, some attempt is made to take a broader view of the preceding detail.*

* Since the date of writing this monograph, there have been some further developments. These have not in any way outdated the theory given in the monograph, but they have provided detailed algebra, and special computer programs, which make the theory applicable to quite complex empirical data. The numerical results in chapters 3 and 5 remain quite satisfactory for illustrative purposes, but see McDonald (1965a, 1965b, and *in press*) for current recommendations as to computing algorithms and computer programs.

CHAPTER 2

GENERAL BACKGROUND

2.0 *Introduction*

Section 2.1 of this chapter is a review of certain aspects of linear factor analysis. Here we are concerned only with those features of the linear model which have implications for the present nonlinear theory. These include the basic equations of principal components analysis, the relations between factor solutions for correlation and covariance matrices, and the problems of estimating communalities and determining the approximate rank of the reduced observation matrix. All of the considerations developed here carry over to the analysis of the nonlinear models.

In sect. 2.2 the need for nonlinear models is discussed, with particular reference to the problem of "difficulty factors." This leads on to an examination in sect. 2.3 of the principles of latent structure analysis (LSA) and of Guttman's radex theory. The major objective here is to show that when the basic principles of LSA are stated in a much more general form than they have usually been given, the linear and nonlinear factor models can readily be derived as particular cases.

It has been claimed by such writers as Guttman and Lazarsfeld that the concepts of linear factor analysis are inappropriate to observations in the form of response categories. The importance of this distinction between "quantitative" and "qualitative" observations, in the context of factor analysis, is implicitly questioned by the arguments of sect. 2.3. In sect. 2.4 an explicit examination is made of the relation between the *observations* made on the real world and the *data* as recorded in a score matrix for the purpose of factor analysis. Here it is shown that multicategory observations can in fact be converted into data suitable for linear factor analysis. The same conversion can be made in the nonlinear models in the sequel. Hence the system being presented here can be taken to apply to multivalued observations, such as test scores; to dichotomous observations, such as test items scored in terms of pass/fail; and also to multicategory items, whether or not the categories of response can be ordered in some way.

In sect. 2.5, the immediate background to the present approach is treated. This involves an examination of Gibson's attempt to develop nonlinear models by an ad hoc adaptation of latent class analysis, followed by a consideration of the analogy to the present problem that can be seen in the methods of curve-fitting by orthogonal polynomials.

The general intention of this chapter is to clear the ground for the theory and practice presented in the sequel. Some of the issues raised here, such as that of "difficulty factors" and the latent structure models, are dealt with in the terms of the present system in the fourth and fifth chapters. Other considerations, such as the procedure for determining the rank of the reduced data matrix, are employed in the sequel without any need for further comment.

2.1 *Linear Factor Analysis*

There is an extensive literature dealing with the basis of the linear factor models, with methods of factoring, and with criteria for testing the fit of the model to observations. A full review of this literature is unnecessary here. The major issues are covered by Holzinger and Harman (1941), Thurstone (1947), Thomson (1950) and Harman (1960). The most rigorous account of the fundamentals is given by Anderson and Rubin (1956). For the present purpose, it will be sufficient to set out the basic equations of the orthogonal factor model only. The question of oblique factor models will be briefly referred to in the final chapter. The object here is to establish the nature of the assumptions in the model and to indicate briefly the current status of the main problems in applying it to observations, insofar as these concern us in the sequel.

In practice, the starting point for a factor analysis is a sample matrix of observations, of order N by n, let us say. In psychology, it typically consists of measurements obtained from each of N subjects on each of n tests, though other applications of the theory can be made. The first problem is to specify the nature of the population with which we are concerned. In relation to "subjects" and "tests" there are at least three models to consider. In what may be called Model I we regard an observation matrix as a sample of size N from an infinite population of subjects performing on n fixed tests. In Model II we regard the observation matrix as a sample of *one member* from an ensemble of conceptual replications involving the same tests and the same individuals. This is analogous to the situation of fitting a functional relationship to a single set of points, as in an individual learning curve, where statements about "error" are concerned with expected fluctuations in performance if this individual were "brainwashed" and his performance repeated. Model II applies in the case of Tucker's (1958; 1960) work on fitting individual functional relationships by factor analysis (cf. sect. 4.8 below). Model I and Model II are carefully distinguished by Anderson and Rubin (1956). Model III is explicitly or implicitly recognized in the work of Guttman (1950; 1955a) at least. In this model, an infinite universe of content, or behavior-domain, is postulated, from which the n "tests" employed are regarded as a sample, though not, perhaps, a random sample. In such a case, the observation matrix might be thought of as a sample both in terms of

persons and tests. However, since there does not seem to be any means whereby we can increase the size of a sample of tests without limit, yet be assured that we are still sampling the same behavior domain, Model III does not seem worth considering. For the same reason, we do not need a converse to Model I, with individuals fixed and tests "random."

To anticipate a little, it is reasonable to consider transposition of the basic factor equations, in the sense of intercorrelating persons instead of tests, in Model II or III but not in Model I. It is generally asserted (cf. Thomson, 1950; Thurstone, 1947) that factor loadings can be calculated, whereas common-factor scores (given nonzero uniquenesses) can only be estimated. Whittle and Wold (cf. Henrysson, 1957, pp. 133–134) point out that errors of estimation of factor scores approach zero as the number of tests increases without limit, just as errors of estimation of factor loadings approach zero as the number of subjects increases without limit. While this theoretical symmetry of the argument can be admitted, it holds only in Model III. For the present purpose it is best to consider Model I only, in which factor scores remain undetermined in the population, as will be seen shortly. Model II will be briefly treated in sect. 4.8.

Consider a random vector

$$y \equiv [y_1 \cdots y_n]$$

whose expected value

$$\mathcal{E}(y) = \bar{y}$$

is not in general a vector of zeros. In the linear, orthogonal, common-factor model we write

$$y - \bar{y} = xA + eU, \tag{2.1.1}$$

where A is an $r \times n$ matrix ($r \leq n$) of rank r, U is an $n \times n$ diagonal matrix,

$$\underline{x} \equiv [x_1 \cdots x_r]$$

and

$$\underline{e} \equiv [e_1 \cdots e_n]$$

are random vectors, with

$$\mathcal{E}(x'x) = I, \ \mathcal{E}(e'e) = I, \ \mathcal{E}(e) = 0, \tag{2.1.2}$$

and

$$\mathcal{E}(e'x) = 0.$$

Since A is of full rank, it follows that $\mathcal{E}(x) = 0$. The vector y can be thought of as a set of measures on n tests, or more generally as a set of observations or *manifest variates*. The vector x is a set of $r \leq n$ unobserved orthogonal

factors, factor scores or *latent variates.* The matrix A represents common-factor *loadings, weights* or *saturations,* analogous to regression weights in multiple regression theory.

Given a sample of N observations of y, the primary object of factor analysis is to estimate an A and U which "account for" the observations. More rarely, estimates of the values of x in the sample are also desired. A wide variety of methods for factoring has been developed (cf. Harman, 1960). Here we consider only the principal components solution, which is employed throughout the sequel.

From (2.1.1) and (2.1.2) we have

$$\mathfrak{R} \equiv \mathcal{E}\{(y - \bar{y})'(y - \bar{y})\} = A'A + U^2. \tag{2.1.3}$$

It should be noted that *strictly* (2.1.2) contains some redundancy. Working straight from (2.1.1), we have

$$\mathcal{E}\{(y - \bar{y})'(y - \bar{y})\} = A'A + \mathcal{E}(Ue'eU) + \mathcal{E}(Ue'xA) + \mathcal{E}(A'x'eU).$$

It is sufficient in everything that follows to assume that

$$\mathcal{E}(Ue'eU) + \mathcal{E}(Ue'xA) + \mathcal{E}(A'x'eU) = U^2,$$

rather than assume that the terms in $e'x$ and $x'e$ vanish individually as in (2.1.2). The point is only worth mentioning because of the fact that while the last equation can be deduced from the principle of local independence in latent structure analysis, the full set of assumptions in (2.1.2) cannot. The question arises in this form in section 2.3. Thus if the model holds, a diagonal matrix U^2 can be found such that the "reduced covariance matrix" $\mathfrak{R} - U^2$ is of rank $r \leq n$. Let M be the $r \times n$ matrix whose rows are the latent vectors of $\mathfrak{R} - U^2$ corresponding to nonzero latent roots, and C be the $r \times r$ diagonal matrix containing the nonzero roots. Then

$$\mathfrak{R} - U^2 = M'CM \tag{2.1.4}$$

and

$$MM' = I. \tag{2.1.5}$$

If we write

$$F \equiv C^{1/2}M, \tag{2.1.6}$$

then the matrix

$$A = LF, \tag{2.1.7}$$

where L is any $r \times r$ orthogonal matrix, is an admissible solution to the problem of determining factor loadings. The matrix F is the principal components solution. Thus far, of course, problems of sampling and of finding the matrix U^2 have been ignored. Equation (2.1.7) represents the "rotation

problem" of linear factor analysis, i.e., the fact that solutions to (2.1.1) can be determined only up to an orthogonal transformation. Psychologists have invoked criteria such as "simple structure" for A to reduce this indeterminacy (Thurstone, 1947). In the sequel, a rotation problem of a rather different kind arises.

Given a principal components solution for the population,

$$y - \bar{y} = xF + eU, \tag{2.1.8}$$

by (2.1.5) and (2.1.6) we have

$$x = (y - \bar{y})F'C^{-1} - eUF'C^{-1}. \tag{2.1.9}$$

Since e is an unobserved random vector, if n is finite and U contains nonzero elements, there is a fundamental indeterminacy in x in the population. It is in this sense, rather than in the sense of sampling theory, that for Model I common-factor loadings can be "calculated" while common-factor scores can only be "estimated." If, as in Model III, n can be allowed to increase without limit, with no alteration of rank, the term $eUF'C^{-1}$ will approach zero. However, as pointed out above, it is difficult to believe that one can ever draw on a "universe of content" or "behavior domain" in this way. If one did operate in terms of Model III, expectations over the poulation of tests could be written down for $(y - \bar{y})(y - \bar{y})'$ as well as for $(y - \bar{y})'(y - \bar{y})$ thus providing a sensible basis for what has been called Q-technique (Cattell, 1952) or P-technique (Burt, 1937). In Model I, this symmetry does not strictly obtain.

In psychological work as a rule, neither the mean nor the variance of a set of observations has any absolute significance. It follows that in finding the matrices M and C required above, the research worker has at least the choice of operating on an obtained covariance matrix, or on the corresponding correlation matrix, or on the (raw) product-moment $\mathcal{E}(y'y)$ without applying the correction for means, as in the procedure given by Eckart and Young (1936). Since in the sequel there are examples in which the first two possibilities are employed, and one case (sect. 4.8) where the Eckart-Young procedure might be considered, it is worthwhile to examine briefly the relations to be expected between results from the three procedures. It will first be shown that *in general* the effect of operating on the matrix $\mathcal{E}(y'y)$ instead of the deviation form is to yield an additional factor, and a rotation of the matrix of factor loadings can be found such that one of its row vectors is simply the vector $\bar{y}$.

We rewrite (2.1.1) as

$$y = (\bar{y} + xA) + eU. \tag{2.1.10}$$

In general the vector $\bar{y}$ will be linearly independent of the row vectors of A. If in such cases we form the matrix A^* by augmenting A with the row vector

$\bar{y}$, i.e.,

$$A^* \equiv \begin{bmatrix} \bar{y}_1 \; \bar{y}_2 & \cdots & \bar{y}_n \\ a_{11} a_{12} & \cdots & a_{1n} \\ \vdots \; \vdots & & \vdots \\ a_{r1} a_{r2} & \cdots & a_{rn} \end{bmatrix}$$

and write

$$\underline{x^*} \equiv [1 x_1 x_2 \cdots x_r],$$

(2.1.10) becomes

$$y = x^* A^* + eU,$$

where

$$\mathcal{E}(x^{*\prime} x^*) = I, \tag{2.1.11}$$

so that

$$\mathcal{E}(y'y) = A^{*\prime} A^* + U^2, \tag{2.1.12}$$

where A^* is of full rank, $(r + 1)$. Thus, *in general*, with subtraction of the same diagonal matrix U^2 the matrix $\mathcal{E}(y'y)$ is of rank $(r + 1)$ given that the covariance matrix is of rank r. The converse obviously holds also. Further, given a factorization B of $\mathcal{E}(y'y) - U^2$, i.e., a matrix B of order $(r + 1) \times n$, and rank $(r + 1)$ such that $\mathcal{E}(y'y) - U^2 = B'B$, there exists an orthogonal matrix L such that

$$LB = A^*, \tag{2.1.13}$$

i.e., such that one of the row vectors of LB is the vector of means, $\bar{y}$. This rather suggests that in general there is nothing to be gained from operating on $\mathcal{E}(y'y)$ as against the covariance matrix, and on the whole the latter would be the preferable starting point, since it is easy to remove the vector of means at the start, instead of letting it remain as a factor.

We consider next the effects of putting the observations in standard form, i.e., with mean zero and variance unity. Writing

$$\underline{K} \equiv \text{diag } \{\text{var } (y_1) \cdots \text{var } (y_n)\},$$

we have

$$z = (y - \bar{y}) K^{-1/2} \tag{2.1.14}$$

for the vector of standardized observations. Then the matrix of intercorrelations

$$\underline{R} \equiv \mathcal{E}(z'z) = K^{-1/2} \mathcal{R} K^{-1/2}. \tag{2.1.15}$$

It follows immediately that if

$$\mathfrak{R} = A'A + U^2,$$

then

$$R = K^{-1/2}A'AK^{-1/2} + K^{-1/2}U^2K^{-1/2}$$

so that if A and U provide a factorization of $\mathfrak{R}$, $AK^{-1/2}$ and $UK^{-1/2}$ provide a factorization of R. However, if we have a principal components solution F, as above, for $\mathfrak{R}$ such that FF' is a diagonal matrix as required by (2.1.5) and (2.1.6), then $FK^{-1/2}$ is not in general the same as the principal components solution for R, but is related to it by an orthogonal transformation. Given the rotational indeterminacy of factor analysis, it is sufficient to regard standardization of the observations as equivalent to multiplying A and U on the right by $K^{-1/2}$. Again, of course, these remarks apply only to population values. In the sequel, the basic operations are carried out on either covariance or correlation matrices according to theoretical considerations that are specific to each case, but the choice is not of fundamental importance. It is true, however, as Anderson and Rubin (1956) show, that sample estimates of the covariance and the correlation matrices may not even yield the same decision as to the rank of the reduced matrix.

We turn now to a brief survey of the current status of problems in dealing with finite samples of observations. The main concerns are the interrelated questions of the rank of the covariance matrix and the determination either of the matrix U^2, or of the *communalities* which are the reduced elements in the leading diagonal after subtraction of U^2. Broadly, following Wrigley (1957), one can distinguish a purely algebraic treatment on the one hand, as in the work of Thurstone (1947) and Guttman (1954a; 1956; 1957a), and a genuinely statistical treatment on the other, stemming from the classical work of Lawley (1940) on maximum likelihood estimation of factor loadings. The second type of treatment is adequately reviewed by Burt (1952), Maxwell (1959), Rao (1955), and Anderson and Rubin (1956). The last two references cited also contain original contributions.

To consider the statistical treatments first, it may be sufficient to note that methods such as that of Lawley (1940) rest on statistical assumptions which cannot readily be justified in the extension of factor methods to nonlinear factor analysis. Hence the theory to be developed here cannot make use of existing sampling theory. Nor has it been possible so far to develop sampling theory appropriate to the nonlinear models. Consequently, practical applications must rest on algebraic criteria for rank such as those of Guttman (1954a). It should be noted that there are still some problems and doubts with respect to the application of the maximum likelihood methods. Thus, Anderson and Rubin indicate that although a decision criterion now exists for testing the hypothesis that the rank r of the reduced covariance matrix

is r_0 against the alternative $r > r_0$, there is still no rational multiple decision procedure if the investigator has no a priori hypothesis as to the rank. This is by far the most usual situation in practice. Wrigley (1958) has indicated other "logical difficulties" with the method. It might even be argued that if a factor is technically significant but trivial in the sense that it corresponds to a very small latent root, it would be sensible in practice to ignore it. This means that instead of asking whether a model such as (2.1.1) of some rank $r \leq n$ fits a given observation matrix, we ask whether a model like (2.1.1) of "low" rank can give a "reasonable" fit to the data, where neither "low" nor "reasonable" can be strictly defined.

The last point leads us to consider "algebraic" treatments of the practical problems. In principle, Thurstone (1947) considered that the diagonal entries of the correlation matrix should be so chosen as to minimize its rank, though none of his practical suggestions could be guaranteed to do this, even "approximately." Albert (1944) has provided an exact (and laborious) algebraic solution for elements of U^2 yielding minimum rank r provided that $r < n/2$, a condition which is most unlikely to hold for exact rank in empirical data, as Harman (1960) and others have pointed out. Guttman (1954a; 1956) has provided theorems on lower bounds to the minimum rank. Kaiser (1961) has extended the proof for one of these. Guttman (1954a) shows that after subtraction of a diagonal matrix U^2 from a correlation matrix R, which leaves $R - U^2$ positive semidefinite, the rank of $R - U^2$ is not less than the number of latent roots of R which are greater than or equal to unity. He also deduces two stronger lower bounds. Ledermann (1937) has supplied an upper bound formula. Wrigley (1948) questions its practical usefulness. More generally, Wrigley points out that more than 20 criteria for determining reduced rank have been suggested ("algebraic" or "statistical") as though it is easier to invent new procedures than to justify existing ones. Much the same is true of the related communality problem.

It must be admitted, then, that a good deal of the practical side of factor analysis is still in a parlous state. It is no part of the present work to add further solutions (or complications) to these problems, but for the practical aspects of the sequel it is necessary to employ one of the existing procedures. For general convenience, Guttman's weak lower bound described above will be taken as the actual criterion for approximate rank in the numerical examples given. It is particularly convenient in the present state of the nonlinear theory to operate in terms of very low rank. As a rationalization, it can be said that such a procedure should yield "low" rank with "reasonable" fit, in the indefinable sense mentioned above.

A brief examination of the assumptions in the orthogonal linear model will serve to conclude this section. These assumptions are completely expressed in (2.1.1) and (2.1.2), and there is little to be gained from restating them in words. Guttman (1954a) has called the set of assumptions in (2.1.2)

the δ-law of deviation, and he has considered the use of other "laws" of which the α, β, and γ are individual parts of (2.1.2), while the ϵ-law is defined on the basis of Model III discussed above. For the present purpose, these alternatives can safely be disregarded. The important points to note are as follows:

Firstly, except for the purpose of making certain tests of significance, no assumptions are made, in respect to the observations y, the factor scores x, or the unique deviations e, about either their distribution functions or their measurability. The distributions, and the variate measures, can be continuous or discontinuous, and in particular we could have the observations y taking the values zero and one only, with specifiable probabilities. This point is important because writers such as Guttman (1950) have claimed that factor analysis is designed only for "quantitative" observations. Presumably this means either continuously measurable observed variables or, at least, variables which can take a "large" number of distinct numerical values. This matter will be considered further shortly.

Secondly, (2.1.2) sets down orthogonality relations only. There is no requirement in the model that the factor scores or unique deviations be mutually independent statistically. It would be possible to have curvilinear relations between them, consistent with zero correlations. This point is of fundamental importance, as it will be shown in Chap. 3 that all of the relations given above for the linear model still hold in nonlinear models. The essence of the present development is in a reinterpretation of the factor scores.

2.2 *The Need for a Nonlinear Model*

A priori, it may seem possible that behavioral data could turn out to be quite persistently linear, or at least that linear models might serve to describe behavioral data to a quite satisfactory approximation (cf. Cattell, 1960). It is well known, after all, that the size of a correlation coefficient does not alter "much" under monotonic transformations of the metric of the variables (cf. Mosteller, 1958). Provided, then, that behavioral variables are monotonically related, a linear model may serve to describe the relations within the usual limits of error in a factor-analytic investigation. Ahmavaara (1957) claims, though without real argument, that nonlinear models are unnecessary. Bargmann (1960), reviewing Ahmavaara's book, implies that nonlinear models can be avoided by some kind of normalizing or scaling of the observations before analysis. He cites as an instance Thurstone's (1947) well-known treatment of the box problem. Thurstone constructed "observations" out of nonlinear but monotonic functions of the three dimensions of a sample of boxes and recovered the dimensions by factor analysis, having first normalized the observations. Provided that the relations between observed variates are monotonic, it seems that one could apply transformations of

metric such that mutual regressions are linear and then proceed to a linear factor analysis. However, this is not possible in general, since there are $n(n-1)/2$ such regressions to consider, for n observed variables. Further, the nonlinear models considered below apply also to observed dichotomies, for which we cannot consider the linearity or nonlinearity of the observed mutual regressions.

Instances can certainly be found in psychological work of curvilinear and *nonmonotone* relations between observations. Eysenck (1958) has drawn attention to the limitations of factorial methods in work on anxiety, where it is well known that certain performance measures are U-shaped functions of anxiety, as expected on the basis of Hull-Spence theory. Another example is the relation between intensity of attitude and position on the attitude scale as studied by Suchman (1950). It might be possible to have measures of intensity and measures of position included in a matrix of observations, in a sufficiently subtle form to prevent them from being recognized as such. In general, though, it can only be said that a great deal of work will have to be done with the nonlinear models before it is established that they are absolutely necessary.

The most persistent problem in the factor-analytic literature which leads us towards nonlinear models might well be called "the mystery of the difficulty factors." Guilford (1941) in a factor analysis of the Seashore Test of Pitch Discrimination obtained a factor which was related to the difficulty of the items. This might be taken to show that different abilities were employed, respectively, in making easy and difficult discriminations in pitch. Ferguson (1941) pointed out that whereas the generally accepted notion of homogeneous or internally consistent tests was one in which the interitem correlation matrix was approximately of rank one, it appeared that a matrix of phi coefficients (product-moment correlations for dichotomies) would be of rank greater than one if the items differed widely in difficulty level. Ferguson's argument was not, perhaps, as rigorous as might be desired. He considered that the explanation lay in the fact that the maximum possible value for phi is a function of the difference in difficulty level of the items, but his discussion also assumed that one had a set of items concerned with the same "content" in which a person passing any item necessarily passed all easier items. This assumption characterizes the "ideal answer pattern" of Walker (1931; 1936; 1940) or "perfect scale," as Guttman (1950) called it. Ferguson also considered the case of correlations between tests of differing difficulty level and noted that these too should give a correlation matrix of "spuriously" high rank. To illustrate, he presented an empirical correlation matrix which may now be recognized as an additive simplex (cf. Guttman, 1954b, and see sect. 2.3, 4.1, and 4.7 below). Wherry and Gaylord (1944) discussing Ferguson's "dilemma," concluded that it is due to the use of the wrong correlation coefficient, and recommended the use of tetrachorics for

factoring dichotomous data. They also showed that in the case of tests of differing difficulty one could expect their mutual regressions to be nonlinear, but contented themselves with recommending in such cases the practical measures of dichotomizing the test scores and again using tetrachorics. As a result, it seems, attention moved from the basic equations of factor analysis, in which the product moment correlation matrix or covariance matrix is only an intermediate step in calculations, so to speak, to the question of finding the best-behaved correlation coefficient (cf. Cattell, 1952, pp. 321–327). Carroll (1945) and Gourlay (1951) indicated conditions under which tetrachoric correlations might yield a "difficulty" factor. It is presumably known, too, that the substitution of coefficients other than product moments in the correlation matrix will yield a matrix which is not necessarily positive semidefinite; hence one can have the embarrassment of finding negative latent roots and nonreal factor loadings. (I have not seen this point specifically discussed in the literature, but workers in the field seem well aware of it.)

At the theoretical level the question has not been taken further, except for claims by Gibson (1959; 1960) that difficulty factors, being due to mutual nonlinear regressions between tests, can be further considered as due to nonlinear regression of tests on factors. This is plausible enough, especially when we consider that the original arguments of Ferguson, Wherry and Gaylord, and Carroll all assumed a perfect scale, though they did not call it that, and one interpretation of the perfect scale (cf. Torgerson, 1958, and see sect. 2.3 below) is in terms of markedly nonlinear relations between items and the latent variate or factor. However, a general account of the way in which difficulty factors result from nonlinear regressions does not seem to have been given yet.

On the empirical side, there is a curious situation. Dingman (1958) compared phi coefficients and tetrachorics and felt that his results constituted a disproof of Ferguson's theory. Comrey and Levonian (1958) obtained as many factors with phi, phi over phi-maximum, and tetrachorics, but obtained anomalous figures with the second of these. They concluded that phi was a much maligned coefficient. The point seems to have been missed, however, that the earlier arguments rested on the assumption that one had scalable data, i.e., data that conform to the perfect scale.

If these two investigations have failed to find difficulty factors, it may simply be the case that the data employed were not remotely scalable. On the other hand, there is any number of scales and quasi-scales to be found in the literature, and any of these if factored in terms of covariances or product-moment correlations (phi coefficients) will yield difficulty factors. Similarly, Guttman (1957b) has listed a number of matrices which, on inspection, are found to have the simplex property, thus resembling Ferguson's early example of difficulty factors in tests as against items. Any of these may

be further evidence in favour of the earlier arguments, though they are open to an alternative interpretation (see below, sect. 4.7). In short, to anticipate a little, the difficulty factors of these early discussions may turn out to be components due to curvilinearity, of which the principal components of the perfect scale (Guttman, 1950) are only a limiting case. One objective of a nonlinear model would be to establish the existence of intermediate cases between the linear model in a single factor, with a correlation matrix of rank one, and the perfect scale, also involving only one factor, with a correlation matrix of high rank.

In order to make these points clearer, however, it is necessary to make a brief examination of the basic principles of latent structure analysis and of Guttman's radex theory. This is done in the section immediately following.

2.3 *Other Relevant Models*

In the more recent period there have been three major developments in the field of models for the "dimensional" analysis of data. These are, respectively, latent structure analysis (LSA) (Lazarsfeld, 1950; 1954; 1960), the radex models of Guttman (1954b; 1955b) and nonmetric factor analysis (Coombs and Kao, 1954; Coombs and Kao, 1955; Torgerson, 1958). The object of this section is to treat only the basic principles of the first two of these developments insofar as they provide a theoretical background to the present approach. A detailed account of some of the models will be given in Chap. 4. Latent structure analysis is relevant because a proper statement of the latent structure principle—the Principle of Local Independence—enables one to derive both the linear and nonlinear factor analysis models as particular cases, as well as a wide range of other models. The radex theory is relevant, since although it contains no direct implications about the factor composition of tests, there are certain interesting nonlinear factor models which entail the radex properties. Nonmetric factor analysis is relevant, since it can be thought of as the development of a particular class of latent structures or nonlinear factor models. Consideration of this last, however has been left to Appendix 3.4A, where certain generalizations on Coombs' conjunctive/disjunctive models are developed.

There is a certain degree of inconsistency and confusion in the basic statements of latent structure analysis. In his primary account of LSA, Lazarsfeld (1950) introduced it in terms of three notions, viz., a unidimensional continuum on which a single unobserved *latent variate* (i.e., factor score) is distributed, a set of dichotomous items or *manifest variates*, and, for each item, a *trace-line* which is a function describing the conditional probability of giving the designated positive response as a function in the single latent variate.

Assuming polynomial functions for the trace-line, Lazarsfeld derived from this an equivalent *latent class* model in which the very notion of a

latent continuum, unidimensional or multidimensional, was dropped out. In this rather restrictive form of LSA, the Principle of Local Independence is as follows:

For a given position on the latent dimension, responses to the items are statistically independent.

A general formulation was given by Anderson (1959), and a closely related treatment was used in a discussion of latent class models (McDonald, 1962a) which was developed in ignorance of Anderson's work. The following is an expanded version of the latter.

Consider a vector of observations, or manifest variates, $y \equiv [y_1 y_2 \cdots y_n]$. They may be continuous or discrete and in particular dichotomous. Let the joint probability density of the manifest variates be denoted by

$$f(y) \equiv f[y_1 y_2 \cdots y_n],$$

taking nonzero values over the manifest domain M. Let there be a latent variate system Ψ whose density function $g(\Psi)$ takes nonzero values over the latent domain L. This latent variate system may be a discrete or continuous latent variate (scalar or vector) characterizing the population of individuals. Let the conditional density function of the manifest variates for given latent charaterization be denoted by

$$h(y \mid \Psi) \equiv h[y_1 y_2 \cdots y_n \mid \Psi].$$

Let the conditional density function of any manifest variate y_i, for given Ψ be denoted by $h_i(y_i \mid \Psi)$. In the following, all integrals are to be understood in the Stieltjes sense, hence they are to be replaced by sums in the discrete case.

In general we may write

$$f(y) = \int_L h(y \mid \psi) g(\psi)\, d\psi. \tag{2.3.1}$$

The general form of the principle of local independence is that

$$h(y \mid \psi) = \prod_{i=1}^{n} h_i(y_i \mid \psi). \tag{2.3.2}$$

In words, for a given latent characterization, the manifest variates are distributed independently. Without any further assumptions, we can write down a generalization on Lazarsfeld's "accounting equations" (Lazarsfeld, 1950, p. 370). Substituting (2.3.2) in (2.3.1) yields

$$f(y) = \int_L \prod_{i=1}^{n} h_i(y_i \mid \psi) g(\psi)\, d\psi. \tag{2.3.3}$$

The expected value of any product or continued product of the manifest variates is a function of the density of ψ and of certain expected values of

the conditional distributions $h_i(y_i \mid \psi)$. Thus, for a manifest variate y_k,

$$\mathcal{E}(y_k) = \int_L \int_M y_k \prod_{i=1}^{n} h_i(y_i \mid \psi) \prod_{i=1}^{n} dy_i \; g(\psi)\, d\psi, \tag{2.3.4}$$

whence

$$\mathcal{E}(y_k) = \int_L \mathcal{E}(y_k \mid \psi) g(\psi)\, d\psi. \tag{2.3.5}$$

Now let $\prod_s y_k$ be the continued product of any stipulated subset consisting of s of the n manifest variates y_k. Accordingly, $\prod_s (y_k \mid \Psi)$ represents the corresponding continued product of the conditional expected values. Then, in the same way,

$$\mathcal{E}(\prod_s y_k) = \int_L \prod_s \mathcal{E}(y_k \mid \Psi) g(\Psi)\, d\Psi. \tag{2.3.6}$$

Equation (2.3.6) is the most general form of Lazarsfeld's "accounting equations" for any kind of manifest variate or latent variate system.

The advantage of this formulation over those of Lazarsfeld is that it is sufficiently general to contain, as particular cases, linear and nonlinear factor analysis models and the *latent profile* model discussed by Gibson (1959) and derived earlier by Green (1952), as well as including the recognized models of LSA proper. We now consider briefly the basis of some of these models.

It should be noted, to begin with, that the function $\mathcal{E}(y_k \mid \Psi)$ is the regression function, as strictly defined, of the manifest variate y_k on the latent characterization Ψ (cf. Kendall, 1951). It can also be thought of as a generalization on Lazarsfeld's *trace-line*. We follow the usual practice for a manifest dichotomy of assigning to it a dummy variate y_k which takes the values zero or unity. Then in such a case

$$p_k \mid \Psi \equiv \mathcal{E}(y_k \mid \Psi) \tag{2.3.7}$$

represents the conditional probability that $y_k = 1$, for given Ψ. If, further, Ψ is a scalar latent variate x, then

$$p_k \mid x = \mathcal{E}(y_k \mid x) \tag{2.3.8}$$

is a function in x which is a trace-line in Lazarsfeld's sense. Thus, we may question the fundamental character of the distinction between dichotomous and "quantitative" observations that has been insisted on by Lazarsfeld and Guttman, as mentioned in the previous section. The basic kinds of model to consider are obtained, rather, by restrictions placed on the latent characterization Ψ, though it can be useful to write down the particular forms for manifest dichotomies separately.

The first special cases to consider are those in which Ψ has a discrete

distribution. If Ψ takes the values Ψ_i with probabilities g_i ($i = 1, 2, \cdots$), (2.3.6) takes the form

$$\mathcal{E}(\prod_s y_k) = \sum_i \prod_s \mathcal{E}(y_k \mid \Psi_i) g_i. \tag{2.3.9}$$

Equivalents to (2.3.9) have been given by Green (1952) and by Gibson (1959). The latter has named this case the *latent profile model*, and he obtains the equivalent to (2.3.9) by way of an extended geometrical proof that

$$[\mathcal{E}(\prod_s y_k)]_i = \prod_s \mathcal{E}(y_k \mid \Psi_i). \tag{2.3.10}$$

Such a proof is unnecessary when we have the principle of local independence in the form of (2.3.2) as our starting point. Moreover, if all the manifest variates are dichotomies, (2.3.9) yields immediately

$$p_s = \sum_i \prod_k (p_k \mid \Psi_i) g_i, \tag{2.3.11}$$

where

$$p_s = \mathcal{E}(\prod_s y_k)$$

represents the probability that simultaneously $y_k = 1$ for s selected y_k. Equation (2.3.11) is the equivalent in the present notation to Lazarsfeld's (1950, p. 370) "accounting equations" for the latent class model for manifest dichotomies. In view of the above, it has been recommended (McDonald, 1962a) that the term *latent class model* be understood to cover the more general case where the manifest variates are "quantitative" or a mixture of "quantitative" and dichotomous variates, as well as the particular case where all members of the set are dichotomies.

The latent class models will be treated further in sect. 4.3 and 5.3. For the present, it is sufficient to note that these models are essentially "nondimensional" in the sense that although the latent classes can be thought of as arbitrary partitions of a unidimensional or multidimensional continuum (Lazarsfeld, 1950; Gibson, 1960), the classes as treated in the model are not ordered or arranged spatially in any way. Although the latent class models can be related to the linear or nonlinear factor model (Green, 1952; Gibson, 1959; Gibson, 1960), they are in a sense the most remote development which we will encounter.

We consider next the group of models in which the latent characterization Ψ is represented as a single latent variate x with a continuous density function. Subvarieties of this group have been identified by writing particular expressions for $p_k \mid x$, the conditional probability of a positive response given the value of the latent variate. Here, in view of the above, these subvarieties can be taken to include any type of manifest variate for which we can write functions $\mathcal{E}(y_k \mid x)$ for their regressions on the latent variate.

Among the models of this group we have

$$\mathcal{E}(y_k \mid x) = \sum_{p=0}^{r} a_{kp}x^p, \tag{2.3.12}$$

the *latent polynomial* (Lazarsfeld, 1950);

$$\mathcal{E}(y_k \mid x) = a_{k0} + a_{k1}x, \tag{2.3.13}$$

the *latent linear model* (Torgerson, 1958);

$$\mathcal{E}(y_k \mid x) = N\{(x - \mu_k)/\sigma_k\}, \tag{2.3.14}$$

where $N(\cdot)$ is the normal ogive function, i.e., the *normal ogive* model (Torgerson, 1958); and

$$\mathcal{E}(y_k \mid x) = \begin{cases} \gamma_k, & x \leq x_k, \\ \beta_k, & x > x_k, \end{cases} \tag{2.3.15}$$

i.e., a "step" or "jump" function, known as the *latent distance* model (Torgerson, 1958).

These models will be treated in the sequel. For the present, it is sufficient to make two points. Firstly, in this more general form, the latent linear model is precisely the single-factor case of the linear factor analysis model, and it is no coincidence that, as Torgerson (1958) notes, the Spearman single-factor formulae can be used to solve for the parameters in this model. Secondly, as Torgerson further points out, in the latent linear model for manifest dichotomies, one factors the covariance matrix formed from the elements $p_{ik} - p_ip_k$ (where p_i is the probability of a "positive" response, and p_{ik} is the probability of joint "positive" responses) to obtain the regression slopes (or trace-line slopes) a_{k1}. More generally, in any of the models for manifest dichotomies, factoring the covariance matrix by the methods given below leads more directly to their parameters as commonly described than if we operate on the product-moment correlation matrix (phi-coefficients). This will be seen later (Chap. 4). However, the choice between covariances and correlations is still not of fundamental significance (cf. sect. 2.1).

Next we consider the group of models in which the latent characterization Ψ requires a multidimensional continuum. The linear factor model of section 2.1 has the property

$$\mathcal{E}(y_k \mid x_0, x_1, x_2, \cdots, x_r) = \sum_{p=0}^{r} a_{kp}x_p \tag{2.3.16}$$

where x_0 is by convention set equal to unity, corresponding to (2.1.10). For manifest dichotomies, this model might well be called the *latent hyperplane,* since it is the obvious extension of the latent linear model (2.3.13). In such a case it is evident that the more general trace-hyperplanes corresponding

to Lazarsfeld's trace-line notion, can be fitted by multiple factor analysis of the covariance matrix, up to an orthogonal transformation. The more general multidimensional model, which provides the basis of the present approach to nonlinear factor analysis, can reasonably be left to the next chapter where it belongs.

One general aspect of this treatment of LSA remains to be considered. We identify a specific LSA model by writing specific functions $f_i(\Psi)$, $(j = 1, \cdots, n)$ in

$$\mathcal{E}(y_i \mid \Psi) = f_i(\Psi).$$

To link this kind of statement with the more usual statements for factor models, we can define a unique deviate e_i by the relation

$$y_i = \mathcal{E}(y_i \mid \Psi) + u_i e_i, \tag{2.3.17}$$

where

$$\mathcal{E}(e_i^2) = 1.$$

Model (2.3.16) for example would then take the familiar form

$$y_i = \sum_{p=0}^{r} a_{ip} x_p + u_i e_i \tag{2.3.18}$$

or, in vector form,

$$y = x^* A^* + eU, \tag{2.3.19}$$

as in (2.1.10). However, the complete set of properties in (2.1.2),

$$\mathcal{E}(e'e) = 1, \qquad \mathcal{E}(x'e) = 0,$$

is not strictly entailed by the principle of local independence. Comparing

$$\mathcal{E}(y'y) = A^{*\prime}A^* + \mathcal{E}(Ue'eU) + \mathcal{E}(Ue'xA^*) + \mathcal{E}(A^{*\prime}x'eU) \tag{2.3.20}$$

obtained from (2.3.19) with

$$\text{Nondiag } \mathcal{E}(y'y) = \text{Nondiag } A^{*\prime}A^*$$

obtained by applying (2.3.6) to (2.3.16), it is clearly implied only that the nondiagonal elements of

$$\mathcal{E}(Ue'eU) + \mathcal{E}(Ue'xA^*) + \mathcal{E}(A^{*\prime}x'eU)$$

be zero. Correspondingly, as pointed out in sect. 2.1, there is redundancy in the conditions (2.1.2) for the development of the factor model. This holds more generally for the nonlinear models to be treated. It would be overpedantic in the extreme to insist on this, hence in the sequel the assumptions (2.1.2) will be employed, purely to simplify the algebra.

We turn now to a brief examination of the principles of Guttman's

radex theory. This can be approached, in a limited way, as a natural development from step-function models like the latent distance model, though this is not the only interpretation. For manifest dichotomies, the latent distance model of (2.3.15) has an interesting limiting case in which we set γ_k equal to zero and β_k equal to unity for all k. (The normal ogive model of (2.3.14) can be considered to approach this limit by letting σ_k approach zero for all k.) Data conforming to this limiting case will form the *ideal answer pattern* of Walker (1931), or *perfect scale* for dichotomies of Guttman (1950). In such a response pattern, there is a unique ordering of the items such that a positive response to any item entails a positive response to all items in one direction from it. For the perfect scale, the correlation (phi) between two items is given by

$$r_{jk} = \sqrt{\frac{p_j}{q_j}\Big/\frac{p_k}{q_k}} = \frac{\alpha_j}{\alpha_k}, \qquad p_j < p_k, \tag{2.3.21}$$

where p_j is the probability of a "positive" response to item j. A correlation matrix in which numbers α_j exist such that

$$r_{jk} = \alpha_j/\alpha_k, \qquad j < k, \qquad j, k = 1, \cdots, n, \tag{2.3.22}$$

for a specified ordering of the variables is called by Guttman (1954b; 1955b) an additive simplex. For the case of quantitative manifest variates, Guttman has developed a theory of the simplex and a further analogue, the *circumplex*, out of the property (2.3.22) rather than from the point of view of the shape of the regression function, as in (2.3.15). In his theory, the regression functions are assumed to be linear and the whole theory is assumed to have quite a different meaning from the mathematically identical theory for manifest dichotomies (Guttman, 1954b, p. 323). In Guttman's factorial treatment of the simplex, the tests are considered to be ordered in respect to their "complexity"—a property which can be represented in terms of sect. 2.1 by stating that the matrix of factor loadings A is triangular, with each "successive" test loading on one more factor than its predecessor. Guttman (1954b; pp. 283–286) explicitly distinguishes the notion of complexity from the notion of difficulty, denying that differences in difficulty level could account for a simplex. This is rather curious, in a way, as we have already noted in sect. 2.2 that Wherry and Gaylord (1944) accounted for difficulty factors in tests, yielding a simplex matrix in Ferguson's (1941) paper, by mutual nonlinear regressions between tests. In fact, their illustration of the form of regression to be expected is a step-function. It seems, then, that we have a second mystery, the "mystery of the simplex factors." Most mysterious of all, the earlier "mystery of the difficulty factors" and the present one may turn out to be, not two mysteries, but one only. Accordingly, one of the tasks to be undertaken (sect. 4.7) is to examine the likely factor models corresponding to Guttman's simplex model. It turns

out that we can and should distinguish single-factor and multiple-factor versions of it.

A further aspect of the simplex "mystery" concerns the theory of the principal components of the perfect scale, given by Guttman (1950; 1954c) and Suchman (1950). Guttman (1954c) has shown that a correlation matrix whose elements are of the form (2.3.22) is of full rank and that the loadings on principal components after the first can be written as "oscillatory" functions in the first principal component. This statement needs qualification in that DuBois (1960) has shown that a diagonal matrix U^2 can be found such that $R - U^2$ is of rank $n/2$. This point will be considered further in section 4.7.

Guttman (1950) gives similar theory for the perfect scale, though operating on a somewhat redundant form of the observation matrix (cf. Lord, 1958). In that case, the principal component scores (which are fully determinate in this model) behave in a similar fashion to the principal component loadings. In view of the preceding, we might be tempted to identify these components with the difficulty factors, "one for each degree of difficulty," conjectured by Ferguson (1941) to explain his simplex-like correlation matrix. However, Guttman and Suchman have offered psychological interpretations of the principal components in the context of attitude measurement, in which the second principal component is identified as intensity of attitude, the third as "involution," and the fourth as "closure" (Guttman, 1954c), on the grounds that separately measured intensity, involution, and closure turn out to be curvilinear functions of attitude position which resemble the "oscillatory" functions obtained for the principal components. Burt (1953), in a long reply to a short note by Guttman (1953) on the relation between the latter's scale theory and the former's treatment of "qualitative" data by factor analysis (Burt, 1950), notes that for a perfect scale the principal component scores are proportional to orthogonal polynomials as tabulated by Fisher and Yates. Burt's discussion is not completely clear, but it seems to be implied that the principal components are orthogonal components of curvilinear regression lines.

In the treatment of the perfect scale given below (sect. 4.6) the principal components are interpreted only as the harmonics of regression lines in the form of a step-function, as in (2.3.15). In the more general simplex model, a single-factor model with step-function regression lines will be distinguished from alternative multiple-factor models.

2.4 *Observations Versus Data*

It has already been noted in the context of linear factor analysis (sect. 2.1) and LSA (sect. 2.3) that we can treat any dichotomous observation as a random variable taking the value unity for a "positive" response and the value zero for a "negative" response, where in general the designation

of the response alternatives as positive and negative is quite arbitrary. Further, if we do so, regression functions $\mathcal{E}(y_i \mid \Psi)$ as discussed earlier can be interpreted as the conditional probability of the response category labelled "positive," given the latent characterization Ψ. In this way, dichotomous data can be treated as a particular case of "quantitative" data. In addition, given $\mathcal{E}(y_i \mid \Psi)$ with this treatment of the response categories, the conditional probability function for the "negative" response category can immediately be written down as the probability complement of the given function. Thus, although the treatment of the response categories is asymmetrical, there is no loss of information and results can be written if desired in a symmetrical, though redundant, form.

In the above treatment of linear factor analysis, and in the extension of the notions of LSA, it has been assumed without careful examination of the point, that our primary observations have been put into a form such that they can be regarded as random variables, in subsequent developments of the models. That is, our starting point is an "observed" vector of *numbers*. Now it is not always obvious how a set of observations on the real world becomes represented by a set of numbers.

A rigorous account will not be given here of the theory of measurement. In the light of discussions by Torgerson (1958) and Coombs (1960) the few distinctions we need may seem nonrigorous and perhaps question-begging. Following Coombs, we distinguish between *observations* and *data*. *Observations* are the events in the real world which we observe; *data* are such representations of the observed events as we subject to analysis. Thus, *one* way of representing dichotomous observations as data is the method of (0, 1) scoring considered above. For all purposes of factor-analytic theory, as dealt with here, it appears sufficient to distinguish two types of data, viz., *measured data* and *category data*.

In *measured data*, a single observation is represented by a single member of the real number system, and in a vector of such data, the elements are mutually independent *functionally*, though not, in general, independent in the probability sense. All discussion up to this point has applied specifically to measured data.

When an observed event consists of the occurrence of one out of $m(\geq 2)$ mutually exclusive and exhaustive events forming a class, we can represent the observation as *category data* by stipulating a fixed order for the m categories and representing this as a vector in which one of the m elements is set equal to unity, corresponding to the occurrence of the corresponding event, and the remaining $m - 1$ elements are set equal to zero. That is, a single observation on one attribute having m categories is represented by a vector of m elements as described. This method of representation has been used by Burt (1950), and an equivalent method is involved in Guttman's (1941; 1950) work on principal component theory. More generally, if we have n

attributes, with m_j categories in attribute j, we can represent a set of observations by a row vector y of $\sum_{j=1}^{n} m_j$ elements, in which each subvector, of m_j elements, consists of a single unity and $m_j - 1$ zeros, ordered as described. Thus the elements in the vector are not functionally independent, in contrast to the case of measured data.

The point of introducing the notion of category data is that it provides for a further extension of linear factor analysis to cover multicategory observations. As Burt (1950) has shown, if we form an $N \times \sum m_j$ matrix Y, from N observations on y as just defined, then $Y'Y$ is a Gramian matrix whose submatrices can be recognized as contingency tables. Burt further shows that it is quite a practical proposition to apply factor-analytic techniques to a sample matrix of the type $Y'Y$. The only point which remains a little unclear in Burt's account of the matter is just what such an analysis means. Let us write

$$y_{jk} = 1, \qquad \text{if event } k \qquad (k = 1, \cdots, m_j) \qquad \text{of event-class } j \text{ occurs,}$$

and

$$y_{jk} = 0, \qquad \text{otherwise.}$$

Then

$$y = [y_{11}y_{12} \cdots y_{1m_1} \cdots y_{n1}y_{n2} \cdots y_{nm_n}].$$

It is then immediately evident that

$$\mathcal{E}(y) = [p_{11}p_{12} \cdots p_{1m_1} \cdots p_{n1} \cdots p_{nm_n}],$$

where p_{jk} represents the probability of occurrence of event k in event-class j, and that in the language of the previous section

$$\mathcal{E}(y \mid \Psi) = [p_{11} \mid \Psi \cdots p_{1m_1} \mid \Psi \cdots p_{n1} \mid \Psi \cdots p_{nm_n} \mid \Psi],$$

where $p_{jk} \mid \Psi$ represents the probability of occurrence of event k in event-class j, for a given latent characterization Ψ. It follows that if multicategory observations are represented as category data, the basic equations of linear factor analysis could be taken to apply, and the interpretation is analogous to that for dichotomies, represented by the (0, 1) scoring system. Because of the relations

$$\sum_{k=1}^{m_j} y_{jk} = 1, \qquad \text{for all} \quad j,$$

there is some redundancy in this representation. In the case of dichotomies, of course, the redundancy is such that the (0, 1) representation as measured data is to be preferred. For the present purpose it is sufficient to establish (i) the dual status of dichotomous observations as measured data or category data, (ii) the fact that linear factor analysis can readily be extended to cover multicategory observations, by representing them as category data.

The further consequence, as will be evident later, is that the system of nonlinear factor analysis to be presented in Chap. 3 is immediately applicable to multivalued or two-valued measured data, and to multicategory observations cast in the form of category data. The system can thus be employed on a wide range of problems.

2.5 *Towards the Present Approach*

In the first section of this chapter we considered the minimum set of assumptions in the linear factor analysis model, the relations between the results of operations on correlation, covariance, and (raw) product-moment matrices, and the outstanding problems of estimation and hypothesis testing in this field. Most of the mathematics given carries over to the nonlinear models treated in the next chapter. In the second section we noted some of the situations in psychology which seem to demand the development of nonlinear factor models. In the last section we have seen how a general statement of the principle of local independence in LSA can lead to a whole family of models according to the restrictive assumptions made as to the distribution of the latent characterization, or the form of the regression function, and we have noted the equivocal status of Guttman's radex models from the point of view of factor analysis or LSA. There was, however, no discussion of methods of analyzing data in terms of these models. There are two further considerations to set down to fill in the background completely. These concern, respectively, explicit attempts to develop a nonlinear factor analysis and some aspects of the general theory of regression.

Gibson (1959) has attempted an ad hoc adaptation of the general latent class model, or latent profile model, to the problem of nonlinear regressions of tests on a single factor or latent variate. In a second paper Gibson (1960) has extended this treatment to the case of two factors. In this treatment an initial analysis of the data yields an estimate of the number of latent classes and an estimated matrix of expected values of the manifest variates within each latent class. In effect, this is a tabulation of the regression function $\mathcal{E}(y \mid \Psi)$ in the present notation for each of a finite set of latent characterizations Ψ. If these latent classes are thought of as obtained by an arbitrary partition of a unidimensional or multidimensional latent continuum, it is possible to make arbitrary assumptions as to the dimensionality of the continuum and the arrangement and spacing of the classes within it. Then plausible nonlinear regressions of the manifest variates on one or more latent variates can be constructed. There is, however, a good deal of indeterminacy in this method as Gibson himself points out, and the indeterminacy would become increasingly worrying as the number of latent classes in the initial analysis becomes large.

There is a close family resemblance between the linear factor model and linear multiple regression. It has been noted earlier that Lazarsfeld's

trace-lines or trace-functions can be interpreted more generally as regression functions of the manifest, or observed, variables on the latent, or unobserved, characterization. This interpretation holds for dichotomous or quantitative manifest variates and for any kind of latent characterization. To consider this point in more detail, we note that the general definition of a regression function, of a dependent variable y on any system of independent variables x, would be

$$\hat{y} = f(x) = \mathcal{E}(y \mid x). \tag{2.5.1}$$

The great majority of psychometric applications of regression theory make use of the linear multiple regression model

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \cdots + a_rx_r. \tag{2.5.2}$$

A notable exception is Saunders' (1956) discussion of a regression function in the form of the quadric

$$\hat{y} = ax_1 + bx_2 + cx_1x_2, \tag{2.5.3}$$

to change his notation slightly. In this model, it is customary to think of one of the independent variables as a *moderator variable* with respect to the parameters of the regression on the other variable. The model itself, however, is symmetrical in the independent variables. A factor analogue of (2.5.3) is treated in Chap. 3. This turns out to be formally identical with the model

$$\hat{y} = ax_1 - cx_1^2 + bx_2 + cx_2^2 \tag{2.5.4}$$

obtained by rotating the quadric surface through forty-five degrees, Maxwell (1957) has considered a regression surface in the form of the general quadric

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2. \tag{2.5.5}$$

A factor analogue of this model arises out of a combination of linear assumptions with Coombs' basic concepts in the conjunctive/disjunctive model (cf. Appendix 3.4A).

More generally, however, the feature of regression theory which points the way to a corresponding development in nonlinear factor analysis is the procedure of fitting single or multiple curvilinear regressions by orthogonal polynomials (cf. Kendall, 1951). In the case of a single independent variable x, a model of the form

$$\hat{y} = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{2.5.6}$$

can be rewritten in the form

$$\hat{y} = c_0 + c_1h_1(x) + c_2h_2(x) + \cdots + c_nh_n(x), \tag{2.5.7}$$

where the functions $h_p(x)$ are orthogonal polynomials in x, i.e.,

$$\mathcal{E}\{h_p(x)h_q(x)\} = 0, \qquad p \neq q. \tag{2.5.8}$$

This immediately suggests a factor analogue in which the $h_p(x)$ are orthogonal functions in a single latent variate. Bartlett (1962) has given examples of factor models of this type and has indicated that functions of the type $h_p(x)$ would behave just like further common factors with regard to the correlation properties of the observations.

These last considerations lead directly to the method of analysis given below. The whole conception of the method rests on a definition of latent variates or factors, such that they are mutually statistically independent, i.e., their joint density function can be written as the product of their separate marginal density functions. The simple fact that if two variates are uncorrelated, or orthogonal, it does not follow that they are statistically independent has not always been recognized by psychometricians. Thurstone (1947) defines statistical independence by orthogonality. If the latent variates are not taken to be statistically independent by definition, *there is no distinction between linear and nonlinear factor analysis*, in the sense used here.

CHAPTER 3

THE THEORY OF NONLINEAR FACTOR ANALYSIS*

3.0 *Introduction*

In sect. 3.1 through 3.3 of this chapter, a formal development is given of the general principles of nonlinear common-factor analysis, up to the point of a formal solution for a model in the form of polynomial regressions of observed variates on a number of latent variates.

In sect. 3.1 we shall see that all the basic relations of orthogonal "linear" common-factor analysis hold under very general conditions. We then establish a general principle for nonlinear models. An approximate statement of this is that a particular nonlinear hypothesis amounts to an assertion that the "factor scores" obtained in ordinary common-factor analysis are distributed on a *curvilinear manifold*, i.e., a curved subspace which spans the factor space.

In sect. 3.2 a complete formal solution is developed for a model in which the regressions of the observed variates on a single latent variate take the form of a polynomial. This can be thought of as a generalization on Spearman's general factor theory. In sect. 3.3, the same treatment is extended further to cover polynomial regressions on a number of latent variates.

A formal development of the more general case where terms involving products of latent variates occur in the regression function does not seem to be feasible at present. In sect. 3.4, this type of case is considered, though somewhat informally, and without any attempt at completeness.

In sect. 3.5, some general comments are offered on the theory of the preceding sections, while sect. 3.6 contains some artificial numerical examples, which serve to illustrate the more practical aspects of the problem.

In sect. 3.7, a scheme is set out for the classification of factor models, with indications of the method for dealing with certain special cases not already covered by the general theory. This provides a suitable background to Chap. 4, where a number of existing models are treated in terms of this theory.

* A part of the theory given in this chapter, illustrated and introduced by means of the quadratic example in sect. 3.6, has been published (McDonald, 1962b) in a somewhat less formal version. See also McDonald (1965a, 1965b, and *in press*) for practical computing procedures.

3.1 *The General Theory*

We consider the model

$$z_j = \sum_{p=1}^{r} a_{jp} Q_p(x_1, \cdots, x_t) + u_j e_j. \tag{3.1.1}$$

The z_j $(j = 1, \cdots, n)$ are a set of n observed variates $(r \leq n)$. It is convenient to suppose that they are in standard form, i.e.,

$$\mathcal{E}(z_j) = 0, \qquad \mathcal{E}(z_j^2) = 1, \qquad j = 1, \cdots, n. \tag{3.1.2}$$

The following theory is not, however, dependent at any point on the second condition in (3.1.2). The x_m $(m = 1, \cdots, t)$ are a set of t latent variates which are statistically independent of each other and of the unique variates e_j, and the e_j also are mutually independent. That is, the joint density function of the x_m and the e_j in the population is given by

$$dF = \prod_{m=1}^{t} f_m(x_m)\, dx_m \prod_{j=1}^{n} g_j(e_j)\, de_j. \tag{3.1.3}$$

We assume without loss of generality that

$$\mathcal{E}(e_j) = 0, \qquad \mathcal{E}(e_j^2) = 1, \qquad j = 1, \cdots, n. \tag{3.1.4}$$

The functions $Q_p(x_1, \cdots, x_t)$, $(t \leq r \leq n)$, are a set of r linearly independent functions of the x_m and a finite number of parameters. We further assume without loss of generality that the $n \times r$ matrix of parameters a_{jp} is of rank r. Given a model of the form (3.1.1), it will always be possible to meet this condition by suitably redefining the Q_p as linear combinations of the initially given functions.

We first show that without loss of generality this model can be reduced to the simpler form of (3.1.10) below, in which the linearly independent functions in the latent variates are orthonormal.

On writing

$$\begin{aligned} z &\equiv [z_1 \cdots z_n], \\ A' &\equiv [a_{jp}], \qquad \text{an } n \times r \text{ matrix of rank } r, \\ e &\equiv [e_1 \cdots e_n], \\ U &\equiv \operatorname{diag}(u_1 \cdots u_n), \\ q &\equiv [Q_1(x_1, \cdots, x_t) \cdots Q_r(x_1, \cdots, x_t)], \end{aligned}$$

(3.1.1) becomes

$$z = qA + eU, \tag{3.1.5}$$

and (3.1.3) and (3.1.4) yield

$$\mathcal{E}(e) = 0, \qquad \mathcal{E}(e'e) = I. \tag{3.1.6}$$

Taking the expected value of $z'z$ over the population, we have, in virtue of (3.1.3) and (3.1.6),

$$R \equiv \mathcal{E}(z'z) = A'\mathcal{E}(q'q)A + U^2. \tag{3.1.7}$$

Thus since A and $\mathcal{E}(q'q)$ are of rank r, $R - U^2$ is of rank r.

Since in the model the elements of $\mathcal{E}(q'q)$ must be finite, to be consistent with (3.1.1) and (3.1.2), there exists an $r \times r$ nonsingular matrix T such that

$$T'\mathcal{E}(q'q)T = I. \tag{3.1.8}$$

On introducing

$$\underline{h} \equiv qT,$$

(3.1.5) becomes

$$z = hT^{-1}A + eU \tag{3.1.9}$$

or

$$z = hB + eU, \tag{3.1.10}$$

where

$$B \equiv T^{-1}A$$

is of rank r. Noting that explicitly

$$\underline{h} \equiv [h_1(x_1, \cdots, x_t) \cdots h_p(x_1, \cdots, x_t) \cdots h_r(x_1, \cdots, x_t)],$$

we see that any model of the form (3.1.1) can be rewritten without loss of generality in the form (3.1.10) where the r elements of the vector h form an orthonormal set of functions in the t variates x, i.e.,

$$\mathcal{E}(h'h) = I. \tag{3.1.11}$$

Accordingly, (3.1.7) becomes

$$R = B'B + U^2. \tag{3.1.12}$$

Let M be the $r \times n$ matrix whose rows are the (normalized) latent vectors of $R - U^2$ corresponding to nonzero latent roots, and C the corresponding $r \times r$ diagonal matrix of nonzero latent roots. Then

$$MM' = I \tag{3.1.13}$$

and

$$M'CM = R - U^2. \tag{3.1.14}$$

It is convenient to write (3.1.13) and (3.1.14) in the form

$$FF' = C \tag{3.1.15}$$

and

$$F'F = R - U^2, \tag{3.1.16}$$

where

$$F \equiv C^{1/2}M$$

is the familiar matrix of principal component factor loadings of $R - U^2$.

An $r \times r$ orthogonal matrix L exists such that

$$LF = B. \tag{3.1.17}$$

Thus it is evident that all the basic relations of linear common-factor analysis, as given in sect. 2.1, obtain also under these more general conditions. In linear factor analysis it is assumed that the r orthonormal functions in $t \leq r$ latent variates x_m can be identified with the latent variates themselves, whereas in this model no such assumption has been made.

We now require some further relations in terms of which we may consider the nature of the orthonormal functions $h_p(x_1, \cdots, x_t)$. The basic relation corresponds to (2.1.9) above.

By (3.1.10) and (3.1.17) we have

$$z = hLF + eU, \tag{3.1.18}$$

whence by (3.1.15)

$$zF'C^{-1} = hL + eUF'C^{-1}. \tag{3.1.19}$$

We rewrite this as

$$v = w + d, \tag{3.1.20}$$

where

$$v \equiv [v_1 \cdots v_r] = zF'C^{-1}, \tag{3.1.21}$$

$$w \equiv [w_1 \cdots w_r] = hL \tag{3.1.22}$$

and

$$d \equiv [d_1 \cdots d_r] = eUF'C^{-1}. \tag{3.1.23}$$

By (3.1.3) and (3.1.11)

$$\mathcal{E}(C^{-1}Fz'zF'C^{-1}) = I + C^{-1}FU^2F'C^{-1} \tag{3.1.24}$$

which we rewrite as

$$\mathcal{E}(v'v) = I + S. \tag{3.1.25}$$

In the principal component analysis of a correlation matrix with unities in the leading diagonal, the product v is the familiar expression for a vector of principal component scores (Thomson, 1950, pp. 363–364). In this con-

text, the elements v_p will be referred to as the *component variates*. It should be noted that when U^2 is nonzero, the component variates are not in general equal to the least-squares estimates of factor scores (Holzinger and Harman, 1941, pp. 267–269), which are given by

$$\hat{x} = zR^{-1}F' \tag{3.1.26}$$

in the linear case. Each component variate v_p consists of a "true" part w_p which is an orthogonal transformation of the orthonormal functions $h_1, \cdots, h_r$, and a deviant part d_p which is a linear function of the unique variates e_i.

In geometrical terms, the vectors v can be thought of as points distributed in r-space. They can further be thought of as having "true" locations w, with random disturbances of location d. Particular hypotheses as to the orthonormal functions h amount to assertions that the "true" components w lie on a curve in the r-space when $t = 1$, or, more generally, on a t-dimensional *curvilinear manifold* when $t > 1$. Corresponding to any specific hypothesis there is a set of $r - t$ equations to the curve or manifold, of the type

$$\Theta_\nu(w) \equiv \Theta_\nu(hL) = 0, \qquad \nu = 1, \cdots, r - t. \tag{3.1.27}$$

It should be noted that here the terms curve and curvilinear manifold include a straight line and a linear space, as special cases.

In its most general form, then, the problem is to discover restrictions of the type (3.1.27) on the distribution of the "true" components w in the r-space. Solutions to this problem as developed in the following sections are subject to three theoretical limitations. (Practical limitations, as yet, are even more severe.) Firstly, it is assumed that all possible linear terms $x_1, \cdots, x_t$ are present in any vector h. Secondly, it is supposed that the remaining orthonormal functions h_p can be reasonably represented by polynomials in the x_m. (Future research could be directed towards an examination of such alternatives as the trigonometric functions of classical Fourier analysis.) Thirdly, it is assumed that the disturbances d_p have a normal joint density function.

Let us examine these assumptions in more detail. We consider hypotheses of the type

$$h = [x_1, \cdots, x_t, ax_1^2 + b, \cdots, cx_m^2 + d, \cdots, gx_1x_2 \cdots], \tag{3.1.28}$$

where t members of the vector h are $x_1, \cdots, x_t$, that is, all possible linear terms are present, but higher degree terms need not all be present.

If the model can be reasonably represented in terms of polynomials, least-squares equations can be written (see below) for the purpose of estimating its parameters, in terms of the moments of the distribution of the "true" components w. These are obtainable, given the moments of the component variates v, *if* we can suppose that the disturbances d have a normal joint density function. This is not a crucial assumption in practice, since these

disturbances are linear functions of the n unique variates e_i. Hence their joint density function should approximate the multivariate normal distribution if the number of manifest variates is sufficiently large, by virtue of the central limit theorem (cf. Kendall, 1951). The covariance matrix S of the disturbances d can be computed from (3.1.25), and their higher moments can be calculated on the basis of their assumed normal distribution. Thus if the higher moments of the v_p are known, the corresponding moments for the w_p can be determined by means of the expansion of

$$\mathcal{E}(v_1^k \cdots v_r^s) = \mathcal{E}\{(w_1 + d_1)^k \cdots (w_r + d_r)^s\}. \tag{3.1.29}$$

It is convenient at this point to develop first the treatment of the single-factor polynomial and then indicate how this is extended to the multiple-factor case.

3.2 *Single-Factor Polynomial*

We consider the particular case in which the h_p are orthonormal functions in a single latent variate x. We adopt the further restriction that each h_p be a polynomial in x of precise degree p. On writing $\mu_0 = 1$ and

$$\mu_p = \mathcal{E}(x^p), \qquad p = 1, 2, \cdots, 2r,$$

the h_p take the explicit form

$$h_p(x) = \frac{\Delta_{px}}{(\Delta_{p-1}\,\Delta_p)^{1/2}}, \tag{3.2.1}$$

where

$$\Delta_p = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{vmatrix}$$

and Δ_{px} is obtained by replacing the last row of the determinant by

$$1, x, x^2, \cdots, x^p$$

(cf. Kendall, 1951; Szëgo, 1959, p. 23).

Without loss of generality we choose an origin and unit for x such that $\mu_1 = 0$, $\mu_2 = 1$, whence

$$h_1(x) = x.$$

We now wish to determine L' such that, in terms of (3.1.22),

$$wL' = [x\ h_2(x) \cdots h_r(x)]. \tag{3.2.2}$$

Introducing the $r \times r$ orthogonal matrix L^*, writing ℓ_p^* for its pth column vector, and $h_p^*(x)$ for an expression in the form of (3.2.1) but containing

coefficients μ_p^* in place of the μ_p, we consider the functions

$$\Phi_p \equiv \mathcal{E}\{w\ell_p^* - h_p^*(w\ell_1^*)\}^2 \geq 0, \qquad p = 2, 3, \cdots, r. \tag{3.2.3}$$

The functions Φ_p are all equal to zero if and only if $L^* = L'$ and $h_p^*(x) = h_p(x)$ for all p. The expression for Φ_p in (3.2.3) is a function in the moments of w, the elements of L^*, and $\mu_2, \cdots, \mu_{2p}$. Given the exact population values for the moments of w, we could substitute them in (3.2.3) and so determine the elements of L^*, and $\mu_2, \cdots, \mu_{2r}$ which make $\Phi_2 = \Phi_3 = \cdots \Phi_r = 0$.

We may now consider the problem of choosing between the alternative hypotheses

$$H_e : h = [x\ h_2(x) \cdots h_r(x)] \tag{3.2.4}$$

and

$$H_0 : h = [x_1\ x_2 \cdots x_r] \tag{3.2.5}$$

on the basis of a finite sample of observations z.

In the absence of a rigorous sampling theory, it seems reasonable to replace the population parameters in the above equations by the corresponding sample values where necessary. Thus from an $N \times n$ matrix Z consisting of N observations of the random vector z we compute

$$\hat{R} = \frac{1}{N} Z'Z,$$

find an estimate of U^2 according to one of the principles referred to in sect. 2.1, and compute F and C by the Hotelling method, say. We then obtain the $N \times r$ matrix

$$V = ZF'C^{-1} \tag{3.2.6}$$

whose N row vectors are independently and identically distributed as the random vector v, and estimate the moments of the joint distribution of the elements of v from the column vectors of V. From $(1/N)V'V$ we obtain an estimate of S using (3.1.25) and estimate the moments of the joint distribution of the disturbances d_p, up to the order required in (3.2.3), using the assumption that their joint density function is normal. These yield in turn estimates of the moments of the w_p by the expansion of (3.1.29). The estimated moments of the w_p can then be substituted in the expansion of (3.2.3).

If we then minimize

$$\sum_{p=2}^{r} \Phi_p$$

with respect to L^* and the μ_p^* we obtain estimates of L' and the μ_p which appear to be optimal in a rather special sense (see sect. 3.5). It should be

noted that when estimates of the type indicated are substituted in the expansion of this function the minimum value can be negative. This reflects the fact that due to sampling errors the estimates of the moments of the w_p obtained need not correspond to the moments of a possible distribution, i.e., the ordinary consistency relations need not hold.

Having determined

$$L^* \simeq L'$$

on the assumption that H_e above is true, we can now test H_e against the alternative H_0, for if the H_e (3.2.4) holds, then by (3.1.19) through (3.1.23),

$$VL \backsimeq H + DL \tag{3.2.7}$$

where H and D are $N \times r$ matrices corresponding to the vectors h and d. It is then possible to test the significance of the prescribed curvilinear relation between elements of the first column of VL^* and those of each succeeding column. If all these prescribed relations reach an acceptable level of significance, then the H_e is confirmed. On the alternative hypothesis, the remaining columns of VL^* should be statistically independent of its first column. If H_e is confirmed, using the explicit form of the function h obtained by minimizing $\sum \Phi_p$, we can then write the complete solution in terms of (3.1.18). If some, but not all, of the curvilinear relations are significant, it will be necessary to consider a further alternative to the models (3.2.4) and (3.2.5). One class of such alternatives can conveniently be treated next.

3.3 *Multiple-Factor Polynomial*

We consider now the model

$$h = [h_1^{(1)}(x_1)\cdots h_{r_1}^{(1)}(x_1)\cdots h_1^{(m)}(x_m)\cdots h_{r_m}^{(m)}(x_m)\cdots h_1^{(t)}(x_t)\cdots h_{r_t}^{(t)}(x_t)], \tag{3.3.1}$$

where $r_1 + \cdots + r_m \cdots + r_t = r$, and the $h_p^{(m)}(\cdot)$ are of the form (3.2.1). We require only an obvious extension of the methods for the single-factor polynomial. As previously, given the moments of the w_p, we wish to find L' such that

$$wL' = h, \tag{3.3.2}$$

where h is now specified by (3.3.1). We introduce $h_p^{(m)*}(x_m)$ for expressions in the form of (3.2.1) but containing coefficients $\mu_p^{(m)*}$ in place of the μ_p, and in the $r \times r$ orthogonal matrix L^* we denote the column vector in a position corresponding to $h_p^{(m)}(x_m)$ in (3.3.1) by $\ell_p^{(m)}$. We then have, corresponding to (3.2.3), the functions

$$\Phi_p^{(m)} \equiv \mathcal{E}\{w\ell_p^{(m)*} - h_p^{(m)*}(w\ell_1^{(m)*})\}^2 \geq 0, \tag{3.3.3}$$

$$p = 2, 3, \cdots, r_m; \qquad m = 1, \cdots, t.$$

As before, the functions $\Phi_p^{(m)}$ are all equal to zero if and only if

$$L^* = L' \quad \text{and} \quad h_p^{(m)*}(x_m) = h_p^{(m)}(x_m) \quad \text{for all } p \text{ and } m.$$

The rest of the treatment for this case parallels the single-factor procedure. Here, we minimize

$$\sum_{m=1}^{t} \sum_{p=2}^{r_m} \Phi_p^{(m)}$$

with respect to L^* and the $\mu_p^{(m)*}$, and after operating on a sample V with L^* as in (3.2.7) we would partition it in conformity with the sections of (3.3.1) corresponding to different x_m, and test the significance of the prescribed curvilinear relations between the elements of column vectors within the submatrices so obtained. If the model is justified, again we obtain a complete solution with (3.1.18), using the explicit form of the functions in h obtained by minimizing $\sum\sum \Phi_p^{(m)}$ (but see sect. 3.5 on the rotation problem).

3.4 *Some Remarks on a Class of More General Cases*

An obvious extension of the work of the last two sections would involve a model containing terms in products of the latent variates. For example, we might wish to consider factor analogues of the quadric regression surfaces

$$\hat{z} = ax_1 + bx_2 + cx_1x_2, \tag{2.5.3}$$

(i.e., Saunders' "moderator variable" case) and

$$\hat{z} = a + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2 \tag{2.5.5}$$

mentioned in sect. 2.5. It is a fairly simple matter to extend the methods just given to cover particular cases of this sort, but a general formal treatment of the complete class of such cases proves rather a forbidding task.

For the present it may suffice if the more general principles required here are merely illustrated by applications to the two cases just mentioned, and to a further quite special case, analogous to a "moderator variable" regression surface

$$\hat{z} = ax_1 + bx_1x_2 \tag{3.4.1}$$

of an unsymmetrical type.

Example 1

The factor analogue of (2.5.3) is

$$h = [x_1 \; x_2 \; x_1x_2] \tag{3.4.2}$$

in the present terminology, since it is evident that the three components in h form an orthonormal set, given that x_1, x_2 are statistically independent and in standard form. It follows from the earlier arguments that such a model

should yield a reduced correlation matrix $R - U^2$ of rank three, and the "true" parts w_1, w_2, w_3, of the three component variates v_1, v_2, v_3, should be related to the h of (3.4.2) by an orthogonal transformation L. Again the problem reduces to one of finding an orthogonal matrix L' such that $wL' = h$. In this case we may define a single function Φ by

$$\Phi = \mathcal{E}\{w\ell_3^* - (w\ell_2^*)(w\ell_1^*)\}^2 \geq 0, \tag{3.4.3}$$

where ℓ_p^* represents, as before, the pth column vector of a 3×3 orthogonal matrix L^*. As before, $\Phi = 0$ if and only if $L^* = L'$. Clearly, the procedure for this case is parallel to the previous ones. That is, we would seek to minimize the sample analogue of Φ as a function of L^*. The only point of difference is in the test of significance which justifies the model. This would amount to a test on the correlation between the elements of the last column of VL^* as previously defined and the products of the pairs of elements in the first two columns of VL^*.

Example 2

A factor analogue to the general quadric regression surface of (2.5.5) is given by

$$h = [x_1 \; h_2(x_1) \; x_2 \; h_2(x_2) \; x_1x_2] \tag{3.4.4}$$

since the five elements in this vector form an orthonormal set. (Thus,

$$\mathcal{E}\{x_1x_2 \cdot h_2(x_1)\} = \mathcal{E}\{x_1h_2(x_1)\} \cdot \mathcal{E}(x_2) = 0$$

and

$$\mathcal{E}(x_1^2x_2^2) = \mathcal{E}(x_1^2) \cdot \mathcal{E}(x_2^2) = 1$$

and the like, with the usual stipulation of origin and unit for each latent variate.) Again the crux of the problem consists of determining an orthogonal matrix L', by introducing analogues of the functions $\Phi_p^{(m)}$ and Φ employed earlier. Here, with the previous notation, we introduce three such functions,

$$\begin{aligned} \Phi_2^{(1)} &= \mathcal{E}\{w\ell_2^* - h_2^{(1)*}(w\ell_1^*)\} \geq 0, \\ \Phi_2^{(2)} &= \mathcal{E}\{w\ell_4^* - h_2^{(2)*}(w\ell_3^*)\} \geq 0, \\ \Phi &= \mathcal{E}\{w\ell_5^* - (w\ell_1^*)(w\ell_3^*)\} \geq 0. \end{aligned} \tag{3.4.5}$$

With analogous operations to those given earlier, this model is justified if in VL^*, the prescribed quadratic relations are found to be significant between the second and first column, and between the fourth and third, and if there is significant correlation between the elements of the last column and products of pairs of elements of the first and third. The importance of this model consists in the fact that it can be shown to arise by combining linear assumptions with the basic principles of Coombs' conjunctive/disjunctive models (see Appendix 3.4A).

Example 3

A factor analogue of the regression surface (3.4.1) is given by

$$h = [x_1 \; x_1x_2], \tag{3.4.6}$$

since

$$\mathcal{E}(x_1^2) = \mathcal{E}(x_1^2x_2^2) = 1$$

and

$$\mathcal{E}(x_1^2x_2) = \mathcal{E}(x_1^2)\,\mathcal{E}(x_2) = 0.$$

It will be noted that this model violates the first assumption made above for the general theory, that all linear terms will be present. It is nevertheless detectable in principle, because of a peculiarity to be expected in the distribution of the component variates in two-space. It is intuitively obvious that there should exist a rotation in the two-space such that

$$\text{var}\,(v\ell_2) \propto (v\ell_1)^2 + \text{const.}$$

since x_2 acts to "moderate" the variability of x_1x_2. The point of general interest about this case is that here we have in fact two latent variates distributed in two-space, without a restriction of the type (3.1.27), as in the general theory, yet detectably different from the two-space linear model. This suggests the existence of a class of models of this type, which may be worth exploring.

More generally, it may be noted that for a set of statistically independent variates $x_1, \cdots, x_t$, one set of multivariate orthonormal polynomials is given simply by all possible products of the orthonormal polynomials for individual latent variates, as given by (3.2.1). That is, if

$$h_{k\ldots s} = \underline{h_k^{(1)}(x_1)} \cdots h_s^{(t)}(x_t), \tag{3.4.7}$$

where $h_p^{(m)}(x_m)$ is given by (3.2.1), then

$$\mathcal{E}(h_{m\ldots s} \cdot h_{u\ldots y}) = 1, \qquad m = u, \cdots, \; s = y, \tag{3.4.8}$$

$$= 0, \qquad \text{otherwise,}$$

i.e., the $h_{m\ldots s}$ form an orthonormal set. This is easily seen by writing

$$\mathcal{E}(h_{m\ldots s} \cdot h_{u\ldots y}) = \mathcal{E}\{h_m^{(1)}(x_1)h_u^{(1)}(x_1)\} \cdots \mathcal{E}\{h_s^{(t)}(x_t)h_y^{(t)}(x_t)\}. \tag{3.4.9}$$

An extension of the earlier theory to the class of cases involving product terms can thus be based on the orthonormal set of functions (3.4.7). However, such extensions seem to demand a somewhat flexible and, in a sense, ad hoc approach. For this reason it is considered sufficient to exemplify this class of cases rather than to treat it exhaustively.

3.5 *General Comments on the Theory*

In order to avoid interruptions to the formal development of the theory, certain general issues which require examination have been suppressed up to this point.

It is convenient at this stage to consider the precise sense in which we are concerned here with *nonlinear* factor analysis. In the analogous treatment of nonlinear regressions by orthogonal polynomials, the procedures employed are still in fact the procedures of *linear* algebra. That is, the functions are nonlinear in the independent variables, but linear in the parameters of the curves. A similar statement holds for the general model (3.1.1), which is nonlinear in the latent variates, but linear in the factor loadings a_{ip}.

In contrast, models such as the normal ogive (2.3.14) and the latent distance model (2.3.15) discussed earlier are nonlinear both in the latent variate and in the parameters of the model. A *general* method of analysis for models of this type does not seem easy to develop. However, it will be seen below that the present procedures can be adapted where necessary to provide a solution for any *prescribed* model of this type. Thus the system as developed is not as restrictive as it might seem from these last remarks.

The process of minimizing the expansions of the functions Φ_p and the like in previous sections resembles a least-squares estimation procedure. We might regard it as providing "quasi-least-squares" estimates. However, it should be emphasized that the process employed rests on the assumption that the "true" parts w of the component variates v lie precisely on the curvilinear manifold to be determined, in which case the minimum value of Φ_p should be precisely zero. In statistical theory to date there is no corresponding least-squares procedure for operating on the observed components v to provide a rotation and best-fitting function, when there is "error" in v, in all directions. In general, when estimates of the joint moments of the w_p as obtained above are substituted in the expansions of these functions Φ_p, there may be a region in which each Φ_p is negative. It could be claimed that any one point in the negative region is as good as any other. However, the procedure above seems, intuitively, to be an optimal one for obtaining point-estimates.

There is a quite distinct sense in which minimizing the function Φ_p can be considered as a least-squares procedure. Suppose that we have the exact values of the joint moments of the "true" points w_p and that the polynomial model under consideration fits precisely. As stated already, each $\Phi_p(\min)$ should then be precisely zero. In contrast, suppose that the orthonormal functions corresponding to the distribution of the latent variates cannot be perfectly represented by a finite set of polynomials. In such a case, we would expect to obtain $\Phi_p(\min) > 0$, and the above procedures would provide best-fitting polynomials to the actual orthonormal functions in the sense of least-squares.

A further point of interest is that no assumptions have been made as

to the distributions of the latent variates, except for the arbitrary choice of an origin and unit such that, for each of them, $\mu_1 = 0$ and $\mu_2 = 1$. The remaining moments of the latent distributions emerge as a by-product of the analysis.

An issue deserving of special consideration is the problem, or rather problems, of rotation in nonlinear factor analysis. One kind of rotation problem has been dealt with, in that the rotation of the "true" points w into the points h with the prescribed functional relations is almost the crux of the theoretical matter. One way to describe this is to say that we seek a rotation of axes in the r-space containing the latent variates, such that the curvilinear manifold on which they lie can be given a simple analytic expression. This rotation may appear to be overdetermined. Thus, to take a simple example, in the model

$$h = [x_1 \; h_2(x_1) \; x_2] \tag{3.5.1}$$

the assertion is that the "true" points w lie on a curvilinear manifold consisting of a parabolic cylinder in general orientation in three-space. There is, in fact, a unique orientation in which the description of the cylinder takes the simple form of (3.5.1). However, given, by the above methods, an analytic solution in this form, which explicitly would be

$$z_i = a_{i1}x_1 + a_{i2}h_2(x_1) + a_{i3}x_2 + u_ie_i, \tag{3.5.2}$$

such a solution can be transformed into any

$$z_i = a_{i1}(c\xi_1 - s\xi_2) + a_{i2}h_2(c\xi_1 - s\xi_2) + a_{i3}(s\xi_1 + c\xi_2) + u_ie_i, \tag{3.5.3}$$

where

$$\begin{aligned} \xi_1 &= cx_1 + sx_2, \\ \xi_2 &= -sx_1 + cx_2, \\ c &= \cos\theta, \\ s &= \sin\theta. \end{aligned}$$

Equation (3.5.3), for any θ, will also constitute solutions for the model, so it is apparent that the ordinary rule of linear factor analysis applies here, that the solution is determined only up to an orthogonal transformation. On the other hand, the description of the cylindrical regression surface in (3.5.2) is obviously much simpler than that in (3.5.3). It rather seems that, in some special cases at least, nonlinear factor analysis has its own "simple structure"—quite different in basis from that of Thurstone. The general rule that emerges here is that the methods given above will yield mathematically simple solutions in terms of some number t of latent variates. In the search for "psychological meaning," a practicing factor analyst is free to apply a $t \times t$ orthogonal transformation to these latent variates,

though in general the mathematical description of the regression surfaces would be more complex after rotation.

Another aspect of rotation is that in the process, product terms in the latent variates appear or vanish almost at will, so to speak. Thus, in the expansion of (3.5.3) above, there will in general be a term in $\xi_1\xi_2$. Conversely, the case of the specialized quadric surface, corresponding to Saunders' "moderator variables," is interesting. Given an initial solution in the form of (3.4.2), if we transform $[x_1, x_2]$ into $[\xi_1, \xi_2]$ as above, with $\theta = 45°$, the model becomes

$$h = \left[\frac{\xi_1 - \xi_2}{\sqrt{2}} \frac{\xi_1 + \xi_2}{\sqrt{2}} \frac{\xi_1^2 - \xi_2^2}{2}\right] \tag{3.5.4}$$

which is a degenerate case of the model

$$h = [x_1\ h_2(x_1)\ x_2\ h_2(x_2)]. \tag{3.5.5}$$

Thus in the factor analogue to the moderator variable case, the interpretation in terms of "moderators" is not determined by the mathematics. More generally, this example serves to show further that there is nothing sacred about the distinction between cases with product terms and cases without them.

Finally for this section, we consider the limitations of the above models. Perhaps the most severe restriction is the assumption generally made above that all linear terms are present. It is an open question whether real data will commonly violate this assumption. The theory is nevertheless very much less restrictive than previously existing theory.

3.6 *Numerical Illustrations*

(i) *Quadratic Case*

The simplest possible case in nonlinear factor analysis would be one in which we wish to choose between the models in which (3.1.10) becomes either

$$z_i = b_{i1}x_1 + b_{i2}x_2 + u_ie_i \tag{3.6.1}$$

or

$$z_i = b_{i1}x + b_{i2}h_2(x) + u_ie_i, \tag{3.6.2}$$

where by (3.2.1) we have

$$h_2(x) = k(x^2 - \mu_3 x - 1) \tag{3.6.3}$$

with

$$k = 1/\sqrt{\mu_4 - \mu_3^2 - 1}. \tag{3.6.4}$$

TABLE 3.6.1

Generating Equations (Raw Score Form)

$$y_i = A_i + B_iX + C_iX^2 + U_iE_i$$

Test	A_i	B_i	C_i	U_i
1	7871	−3425	400	75
2	5982	−2680	300	94
3	0	100	0	10
4	−1934	880	−100	73
5	−6117	2710	−300	87
6	0	200	0	100
7	−8096	3600	−400	226

As a numerical illustration of this case, observations on seven "variables" were constructed, using a table of random numbers (zero through nine) in rectangular distribution for the latent variate x and the unique variates, with arbitrarily chosen parameters in the quadratic equation. These generating equations are given in "raw score" form in Table 3.6.1. The equations themselves can be transformed to standard score form, using the moments of a rectangular distribution of the numbers zero to nine ($\mu_1 = 4.5$, $\mu_2 = 28.5$, $\mu_3 = 202.5$, $\mu_4 = 1533.3$). The transformed equations, which correspond to (3.6.2) are given in Table 3.6.2. A score matrix consisting of one hundred "individuals" by seven variables was constructed using the equations of Table 3.6.1. This matrix was processed by a computer to obtain the correlation matrix R, the matrix of principal-component factor loadings F, and the latent roots C as shown in Tables 3.6.3 and 3.6.4. Iterative estimates of communalities on the hypothesis of rank two were used to obtain the reduced correlation matrix.

TABLE 3.6.2

Generating Equations (Standard Score Form)

$$z_j = b_jx + c_j(x^2 - 1) + u_je_j$$

Test	b_j	c_j	u_j
1	.170	1.115	.073
2	.026	1.126	.123
3	.995	.000	.099
4	−.076	−1.087	.276
5	.013	−1.128	.114
6	.894	.000	.447
7	.000	−1.108	.218

TABLE 3.6.3

Correlation Matrix

(Communalities $1 - u_j^2$ in Leading Diagonal)

	1	2	3	4	5	6	7
1	(.990)						
2	.987	(.983)					
3	.056	.027	(.933)				
4	−.958	−.957	−.083	(.935)			
5	−.990	−.986	−.000	.961	(.991)		
6	−.032	−.062	.897	.013	.086	(.869)	
7	−.971	−.967	.044	.941	.972	.106	(.957)

Not surprisingly, two factors adequately account for the correlation matrix. The 100×2 matrix of component variates $V = ZFC^{-1}$ provided the scatter diagram of Fig. 3.6.1. Because of the large communalities and correspondingly small random deviations, a U-shaped curve is evident by inspection in this example. It is easy to see that the angle of rotation θ is close to 90°, k (the distance from the origin to the vertex) is close to unity, and μ_3 (the perpendicular distance from the origin to the axis of the parabola) is close to zero.

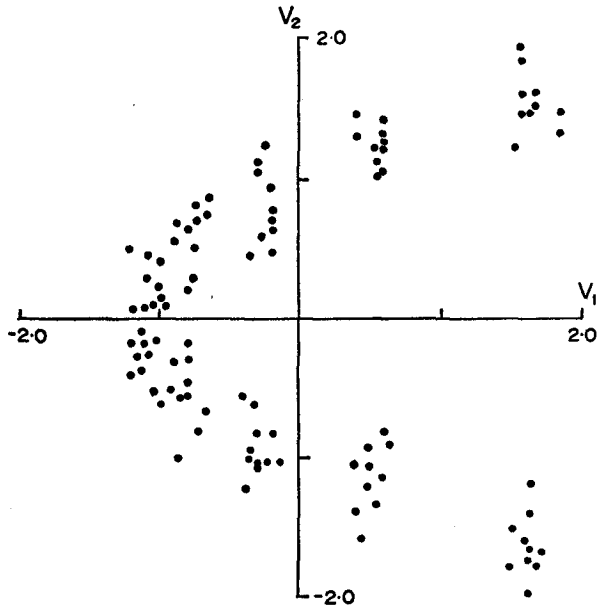
The estimated covariance matrix (S of (3.1.25)) of the disturbances $[d_1\ d_2]$ of the points $[v_1\ v_2]$ from their "true positions $[w_1\ w_2]$ on the suspected parabola is given by

$$\begin{bmatrix} .0176 & -.0017 \\ -.0017 & .0504 \end{bmatrix}.$$

TABLE 3.6.4

Factor Matrix and Latent Roots

	I	II	u_j
1	.994	.042	.098
2	.991	.011	.132
3	.014	.966	.258
4	−.965	−.065	.255
5	−.995	.015	.092
6	−.071	.930	.361
7	−.977	.050	.207
C = diag	[.4853	1.807]	



Scatter diagram of component variates.

Figure 3.6.1

TABLE 3.6.5

Estimated Moments of

(v_1, v_2), (d_1, d_2) and (w_1, w_2)

Moment	v	d	w
m_{20}	1.0176	.0176	1.0000
m_{11}	−.0017	−.0017	.0000
m_{02}	1.0504	.0504	1.0000
m_{30}	.5694	.0000	.5694
m_{21}	−.0235	.0000	−.0235
m_{12}	.9762	.0000	.8762
m_{03}	−.0452	.0000	−.0452
m_{40}	1.9576	.0009	1.8511
m_{31}	−.0272	−.0004	−.0217
m_{22}	1.5663	.0009	1.4974
m_{13}	−.0981	−.0019	−.0910
m_{04}	2.0684	.0076	1.7584

The estimated joint moments of v_1, v_2 are given in Table 3.6.5, together with the moments of d_1, d_2, calculated on the assumption that they have a normal bivariate density function, and also the moments of w_1, w_2, estimated from the preceding on the basis of equation (3.1.29).

Substituting the estimated moments of w_1, w_2 in the expansion of Φ_2 as given by (3.2.3) we minimize Φ_2 with respect to θ^*. We thereby obtain

$$\theta^* = 89°, \qquad \mu_3^* = 0.00077, \qquad k^* = 1.1618;$$

whence also $\mu_4^* = 1.741$ by an analogue of (3.6.4). At these values, $\Phi_2 = -0.017$. Since this is a constructed example, we can compare the obtained values for the third and fourth moments of x with the true values $\mu_3 = 0$ and $\mu_4 = 1.8$ for a rectangular distribution in standard form.

At this point the necessary information has been obtained for reconstructing the quadratic functions required by the model (3.6.2). To do this, we operate on the matrix of principal-component factor loadings given in Table 3.6.4 with the estimated orthogonal matrix L', corresponding to a rotation of axes through $\theta = 89°$, to give the coefficients $[b_{j1}\ b_{j2}]$. This estimate of L' is

$$L' = \begin{bmatrix} .0175 & .9998 \\ -.9998 & .0175 \end{bmatrix}.$$

The corresponding estimate of h_2 is

$$h_2(x) = 1.1618(x^2 - .00077x + 1).$$

Substituting in (3.6.2) this estimated function yields the quadratic expressions given in Table 3.6.6. These may be compared with the generating equations given in Table 3.6.2.

However, in a real case, it would be essential first to justify the choice of the model (3.6.2) as against the model (3.6.1). In principle, this is done

TABLE 3.6.6

Estimated Equations

$$z_j = b_j x + c_j(x^2 - 1) + u_j e_j$$

Test	b_j	c_j	u_j
1	.059	1.154	.098
2	.028	1.151	.132
3	.966	−.003	.258
4	−.081	−1.119	.255
5	−.001	−1.157	.092
6	.928	−.101	.361
7	−.034	−1.136	.207

TABLE 3.6.7

Analysis of Variance:

Regression of v_1 on v_2

Source	df	SS	MS	F
Linear regression	1	0.0152	0.0152	<1
Quadratic component	1	79.9731	79.9731	356.31***
Residual	97	21.7717	0.2244	
Total	99	101.7600		

by operating on the component variates $[v_1\ v_2]$ with L' before carrying out the test of significance. Since in this example the transformation differs only trivially from an interchange of axes, it was considered sufficient to calculate the correlation ratio of v_1 on v_2 and test its significance.

This procedure gave a correlation ratio of 0.88 with an F of 45.66 on 7 and 92 d.f., using eight class intervals on the v_2 axis. An analysis of variance of the regression of v_1 on v_2 is given in Table 3.6.7. This confirms the hypothesis.

(ii) *Cubic Case**

The preceding example served to illustrate the essential features of the present approach, but in some ways it is deceptively simple. As soon as we consider a case involving more than two dimensions, and requiring a considerable amount of rotation, it becomes evident that although the above theory provides a complete solution in principle for a variety of cases, its practical application will require either the use of approximate methods or the development of computer programs for the precise solution of most problems.

In order to show how rapidly the practical problems increase as we move to other cases, a numerical example was constructed in which the regression functions are of the third degree. To facilitate computations, this example is free from unique variations and consists of five "observations" only on six variables. Five "observations" are sufficient for this case. More would be required, of course, in an empirical version of it to test the model adequately.

Taking the five values in Table 3.6.8 for the latent variate x in standard form, we obtain the expressions for $h_2(x)$ and $h_3(x)$ that make up an orthonormal set with x, viz.,

$$h_2(x) = \sqrt{\frac{10}{7}}\,(x^2 - 1)$$

* While the numerical results given here are worth reporting, the computational method employed is now superseded by methods described by McDonald (1965a).

and

$$h_3(x) = \frac{5}{3}\left(x^3 - \frac{17}{10}x\right).$$

The values taken by these expressions, for the given values of x, are also given in Table 3.6.8.

TABLE 3.6.8

$H = h(x)$ for given x

h_1	h_2	h_3
−1.414	1.195	−.707
−.707	−.597	1.414
0	−1.195	0
.707	−.597	−1.414
1.414	1.195	.707

TABLE 3.6.9

B (Transposed)

1	0	0
0	1	0
0	0	1
.707	.707	0
.707	0	.707
0	.707	.707

TABLE 3.6.10

$Z = HB$

−1.414	1.195	−.707	−.155	−1.500	.345
−.707	−.597	1.414	−.922	.500	.578
0	−1.195	0	−.845	0	−.845
.707	−.597	−1.414	.078	−.500	−1.422
1.414	1.195	.707	1.845	1.500	1.345

TABLE 3.6.11

$R = Z'Z$

(1)					
0	(1)				
0	0	(1)			
.707	.707	0	(1)		
.707	0	.707	.5	(1)	
0	.707	.707	.5	.5	(1)

TABLE 3.6.12

F

.577	.816	0
.577	−.408	.707
.577	−.408	−.707
.816	.289	.5
.816	.289	−.5
.816	−.577	0

C = diag [3.0 1.5 1.5]

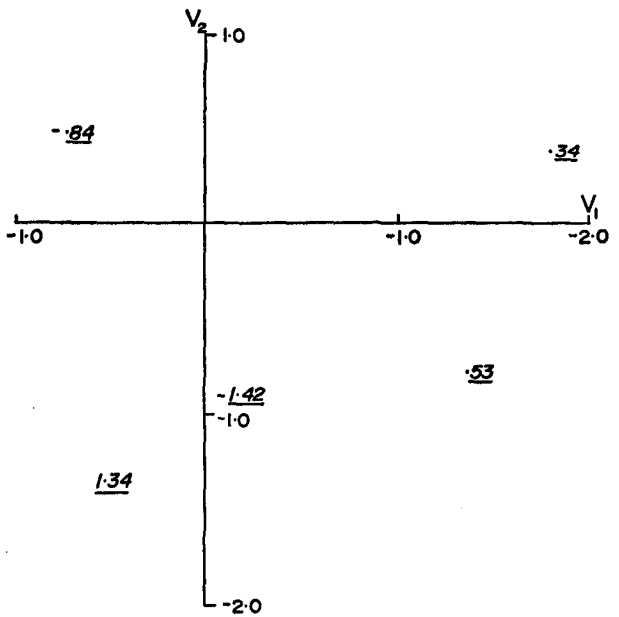
TABLE 3.6.13

$V = ZFC^{-1}$

−.535	−1.354	1.345
.063	−.911	−1.422
−.690	.488	−.845
−.752	1.400	.577
1.915	.378	.345

A matrix of factor loadings B, given in Table 3.6.9, was chosen arbitrarily, so as to require considerable rotation from the corresponding principal-components matrix, in order to recover it in the analysis. The matrix product $HB = Z$ is the score matrix on which the subsequent analysis is based. This is given in Table 3.6.10. The score matrix Z yields in turn the correlation matrix R in Table 3.6.11, the matrix F of principal-component factor loadings, and the diagonal matrix C of latent roots given in Table 3.6.12. From these we obtain the matrix of component variates $V = ZFC^{-1}$ given in Table 3.6.13. A three-dimensional scatter diagram of the component variates v_1, v_2, v_3 is given in Fig. 3.6.2. Inspection of the graph does not reveal any obvious curve on which the points lie in the three-dimensional space.

The expansions for the functions Φ_p of (3.2.3) are, in fact, too cumbersome to use in any case more complicated than the simple quadratic. For the purposes of the present example it was intended to use an approximation formula for small rotations. An initial graphical rotation is therefore required before the approximation formula can be used. The required initial rotation is not at all obvious from inspection of Fig. 3.6.2, so a wire model was con-



Scatter diagram of V_2 on V_1 with values of V_3 given at the points of the graph.

Figure 3.6.2

TABLE 3.6.14

$L^{(1)}$

$$\begin{bmatrix} .40 & .60 & .75 \\ .90 & -.25 & -.25 \\ .10 & .75 & -.60 \end{bmatrix}$$

TABLE 3.6.15

$ZFC^{-1}L^{(1)}$

$$\begin{bmatrix} -1.29 & 1.02 & -.87 \\ -.94 & -.80 & 1.12 \\ .09 & -1.17 & -.14 \\ 1.02 & -.34 & -1.26 \\ 1.14 & 1.31 & 1.13 \end{bmatrix}$$

structed on axes in three dimensions. This gave a twisted curve in the space, which could be seen as a quadratic parabola from one viewing angle and as a cubic curve from another angle. By rough observation the coordinates of these lines of sight with reference to the given axes could be made to yield the orthogonal matrix $L^{(1)}$ given in Table 3.6.14. This provides the initial rotation required.

Applying the rotation $L^{(1)}$ to V yields the product $V^{(1)} = ZFC^{-1}L^{(1)}$ given in Table 3.6.15. In Fig. 3.6.3, the second and third columns of $V^{(1)}$ are plotted against the first. The form of the functions is now readily perceived.

Calculating the required joint moments of $v_1^{(1)}$, $v_2^{(1)}$, $v_3^{(1)}$ and substituting these in the small angle approximation formula for Φ_2 yields, on minimizing this function, an increment of rotation ΔL given in Table 3.6.16. Since ΔL

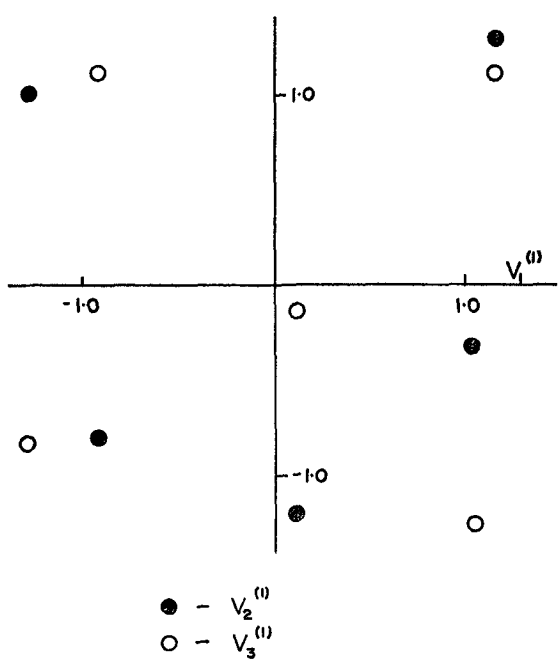


Figure 3.6.3

TABLE 3.6.16

ΔL

1	.024	−.267
−.024	1	.243
.267	−.243	1

TABLE 3.6.17

L_d

.966	.023	−.251
−.023	.971	.229
.258	−.236	.941

TABLE 3.6.18

$ZFC^{-1}L_1L_d$

−1.494	1.166	−.261
−.601	−1.063	1.107
.078	−1.101	−.422
.668	−.009	−1.520
1.363	1.032	1.077

is not an orthogonal matrix, it was normalized by columns to remove some of the scale distortion, yielding the matrix L_d given in Table 3.6.17. Applying L_d to $V^{(1)}$ yields the matrix $V^{(2)} = ZFC^{-1}L^{(1)}L_d$ given in Table 3.6.18.

The values of Φ_2 were, respectively, 0.911 with zero rotation, 0.177 after applying $L^{(1)}$, and 0.095 after applying $L^{(1)}L_d$. With further iteration, since the example is free from unique variation, Φ_2 could be made to approach zero, and the rotated values of V would approach the values in the matrix H given in Table 3.6.8. (If we cheat a little, by operating on V with a rotation matrix L that is given by $F'B$, then $VL = H$ and Φ_2 is equal to zero.) If we stop the analysis at this point, the obtained specification equations are as given in Table 3.6.19. These may be compared with the true specification equations, also given in Table 3.6.19. While further iteration would be desirable, the approximation is not bad.

TABLE 3.6.19

Specification Equations $z_j = a_0 + a_1x + a_2x^2 + a_3x^3$

Theoretical				Obtained			
a_0	a_1	a_2	a_3	a_0	a_1	a_2	a_3
0	1	0	0	−.114	.993	.118	.010
−1.195	0	1.195	0	−1.096	−.826	1.199	.557
0	−2.833	0	1.667	.002	−2.607	.219	1.551
−1.690	1.414	1.690	0	−.857	.119	.932	.400
0	−2.592	0	2.357	−.080	−1.140	.239	1.075
−1.690	−4.006	1.690	2.357	−.774	−2.425	1.003	1.461

Discussion

The above two examples serve to show that the theory can already be applied to simple numerical cases.

An important point emerging from the second example is that visual inspection of bivariate marginals of the joint distribution of the component variates need not, necessarily, yield a preliminary hypothesis as to the model which is most appropriate to the given data. In any empirical case of more than two dimensions, a considerable amount of rotation may be required, perhaps on the basis of pure trial and error, before an appropriate hypothesis can be formulated. This point underlines the importance of developing computer programs to make the rotation problem tractable.

In the empirical work presented in Chap. 5, approximate methods of various kinds have been used where necessary. The reason for this will by now be evident. A good deal of further work will be necessary before the above methods can be applied to empirical data as a matter of routine, even though the problems involved are now solved in principle.

3.7 *The Family of Factor Models*

A reasonably comprehensive classification of factor models can be set out on the basis of the following:

(i) The measures on the observed variates,
(ii) The measures on the latent variates,
(iii) The form of the regression function of observed variates on latent variates.

These will be considered in turn.

(i) *Observed Variates*

As pointed out earlier, various distinctions have been recognized among types of observation or data in factor analysis and LSA. Guttman and Lazarsfeld speak of "quantitative" versus "qualitative"; Green (1952) contrasts dichotomous with "continuous" variables. Here, it seems sufficient, as indicated in sect. 2.4, to distinguish *measured data* and category data as there defined. As a matter of convenience, in Chap. 4, some theoretical developments will be given for measured data only (which, it will be recalled, includes dichotomous data scored 1 for "positive" and 0 for "negative" response). In other cases we will deal specifically with dichotomies. The generalization to multivalued measured data is fairly obvious in these cases.

(ii) *Latent Variates*

Here we may fairly safely follow the conventional distinction between models based on latent classes and models based on latent continua. As

mentioned earlier, Lazarsfeld (1950) and Gibson (1959; 1960) have treated latent classes as arbitrary partitions of a latent continuum. A question of interest, then, is whether a "genuine" discontinuity in the values taken by a latent variate is detectable in practice and/or in principle. This question is taken up in sect. 4.3.

(iii) *Regression Function*

The theory as developed in this chapter has been restricted to regression functions in the form of polynomials, because of their usefulness in leading to expressions in the moments of the latent distribution. The important distinction here, then, is between polynomial regression functions and the rest. There is, however, a method available for transferring from a polynomial model to a prescribed alternative, by performing a harmonic analysis of the alternative in terms of orthogonal polynomials. The general principles of this procedure are as follows:

Let

$$\varphi_j(x) = \mathcal{E}(z_j \mid x) \qquad (j = 1, \cdots, n)$$

be prescribed regression functions of the observed variates z_j on a single latent variate x whose density function is $g(x)$, for which the polynomials $h_\nu(x)$ form an orthonormal set. To the function $\varphi_j(x)$ let there correspond the formal Fourier expansions

$$\varphi_j(x) \sim c_{j0}h_0(x) + c_{j1}h_1(x) + \cdots + c_{j\nu}h_\nu(x) + \cdots \tag{3.7.1}$$

where the coefficients $c_{j\nu}$ (the Fourier coefficients of $\varphi_j(x)$ with respect to the given set of $h_\nu(x)$), are defined by

$$c_{j\nu} = \int \varphi_j(x)h_\nu(x)g(x)\,dx. \tag{3.7.2}$$

Then (cf., Szëgo, 1959), every partial sum of the series (3.7.1) has the following minimum property: If we write

$$\Psi_j(x) = \alpha_{j0}h_0(x) + \alpha_{j1}h_1(x) + \cdots + \alpha_{j\nu}h_\nu(x) + \cdots + \alpha_{jr}h_r(x),$$

where $r \geq 0$ is a fixed integer and the $\alpha_{j\nu}$ are variable coefficients, the integral

$$\int \{\varphi_j(x) - \Psi_j(x)\}^2 g(x)\,dx \tag{3.7.3}$$

becomes a minimum if and only if $\alpha_{j\nu} = c_{j\nu}$ for all p. That is, the Fourier coefficients of a prescribed function $\varphi_j(x)$, in the sense of (3.7.2), will be the coefficients of an orthonormal series, any finite section of which provides a best fit to the prescribed function in the sense of least-squares.

Provided that an expression can be written down for the latent density function, an initial solution in terms of polynomials can then be converted

on demand to a solution in terms of the parameters of an alternative model, using these principles. In sect. 4.4 and 4.5, this procedure is applied to the normal ogive model and the latent distance model.

The general theory as developed in sect. 3.1 is regarded as supplying a unified theory of factor and latent structure models. The practical approach in terms of polynomials supplies approximations to models other than polynomials through the first few terms of their Fourier series.

CHAPTER 4

SPECIALIZATIONS OF THE THEORY

4.0 *Introduction*

In this chapter, the general theory of nonlinear factor analysis is shown to yield a series of special developments which supply a treatment of existing problems or an alternative analysis for existing models.

In sect. 4.1 the problem of *difficulty factors*, mentioned above in sect. 2.2, is revisited. Here it is shown that a difficulty factor could be expected to result from the curvilinear regressions of easy and difficult tests on a general factor of ability. This is a hypothesis that can be confirmed by the general methods of nonlinear factor analysis.

In sect. 4.2 it is shown that the single-factor polynomial model treated in sect. 3.2 becomes the latent polynomial model discussed by Lazarsfeld, in the case where all the observed variables are dichotomies.

In sect. 4.3, an examination is made of the conditions under which a discrete distribution of a latent variate could be detected by the present methods. The discussion in this section is restricted for convenience to a very simple case where we might be concerned to choose between a latent linear model and a latent dichotomy to account for a set of observations.

In sect. 4.4 and 4.5 it is shown that by a harmonic analysis of the functions in the normal ogive and the latent distance models, it is possible to estimate the parameters of these on the basis of a preliminary analysis of a given set of observations in terms of polynomials. The perfect scale is then treated very briefly (sect. 4.6) as a common limiting case of these two models.

Some limited aspects of Guttman's radex theory are taken up in sect. 4.7. Here it is shown that the simplex property as discussed by Guttman is consistent with a nonlinear factor model as well as with the notion, which Guttman appears to favour, of an ordering of the tests in terms of their factorial complexity. It follows that it would be strongly advisable to apply the present methods of nonlinear analysis to any data which possess an approximation to the simplex property before accepting the interpretation in terms of complexity. In this section a limited treatment is also given of a class of nonlinear models in which the regression functions for each observed variable are obtained by translating a prescribed function along the axis representing a single latent variate. The perfect scale, and a restricted form of the normal ogive model, are particular cases of this class.

Finally for this chapter we describe some extensions, based on the general

theory of nonlinear factor analysis, to Tucker's treatment of the fitting of a set of functional relations by linear factor analysis. Alternative treatments are compared and contrasted. One of these is based on Tucker's work and the methods of Chap. 3, while the other is based on the classical procedures for fitting curves, using orthogonal polynomials.

The contents of this chapter provide a broad sampling of the domain of possible specializations of the theory. These possibilities have not yet been examined, or even listed, exhaustively. At the same time, five of the eight theoretical sections in this chapter provide the necessary background for the empirical examples in Chap. 5. Each of the empirical illustrations to be given belongs with a particular section of this chapter.

Sections 4.1, 4.2, 4.3, 4.5, and 4.8, are respectively, illustrated by sections 5.1, 5.2, 5.3, 5.5, and 5.8, hence the former can be read in conjunction with the latter, instead of taking the theoretical and empirical treatments separately.

4.1 *Difficulty Factors Revisited*

The notion of *difficulty factors* was discussed in sect. 2.2. This is the notion that a set of items or tests which differ "widely" in difficulty level *may* yield a reduced covariance matrix of rank greater than unity, even when the set is "really" unidimensional or homogeneous in some sense. Following Gibson (1959), it was noted that a plausible interpretation of such an occurrence is in terms of nonlinear regressions of the tests or items on the latent variate or factor. The argument is not at all rigorous, and it would be quite possible in principle to have linear regressions when the observed variates have "widely" spaced means. In the case of items, treated as measured data, the item mean is of course the proportion of "positive" responses. In the case of a battery of tests, the tests must have a common metric for the argument to have meaning.

A *difficulty factor* is detectable in principle by the process of establishing a functional relation between the test or item means and their loadings on the suspect factor. The real problem is to show that the items or tests are in fact "homogeneous" or "unidimensional" in some acceptable sense. Clearly, linear factor analysis cannot serve this purpose, and it remains a possibility that some difficulty factors represent genuine inhomogeneities in the tasks, in that the easy and difficult tasks may after all be different in kind as well as in difficulty. (The factorization by Guilford (1941) of the Seashore Tests of Pitch Discrimination may be a case of this sort.)

Let us approach the question from a different angle. Suppose we have a set of data which is reasonably fitted by the second-degree case of the single-factor polynomial of sect. 3.2, i.e., (3.1.10) yields

$$y_i = b_{i0} + b_{i1}x + b_{i2}\frac{x^2 - \mu_3 x - 1}{\sqrt{\mu_4 - \mu_3^2 - 1}} + u_i e_i, \tag{4.1.1}$$

where

$$\mathcal{E}(y_i) = b_{i0},$$

is not necessarily zero. In a quite precise sense, the manifest variates y_i form a unidimensional or "homogeneous" set, yet the reduced covariance matrix is of rank two. The family of curves representing the regressions of the observed variates on the latent variate could be such as is shown in Fig. 4.1.1. *If further* the magnitudes of the coefficients b_{i2} correlated highly with the means of the manifest variates, this would correspond to a progressive change in curvature of the regression functions, as shown in the figure, from the "easy" items or tests through to the "difficult" ones.

These considerations provide one application of the methods of Chap. 3. In the first empirical example of sect. 5.1, it is first shown that a difficulty factor emerges, in that the loadings of the tests on one factor are a linear function of the test means. It is then shown that this factor corresponds to the quadratic term in the model (4.1.1), hence that in this case the difficulty factor corresponds to variations in curvature of the regressions of the tests on a single factor.

4.2 *The Latent Polynomial*

In the last section we considered an application of nonlinear factor analysis to a problem which may involve multivalued observed variates. In this and the next three sections we examine the application of the theory to four existing factor models, the latent polynomial, the latent class model, the normal ogive and the latent distance model. The common feature of these models is that they have all been recognized as particular cases in latent structure analysis. With the exception of the normal ogive, they have been developed by Lazarsfeld (1950) within the framework of LSA.

As mentioned earlier, in the latent polynomial model, a single latent variate is assumed, and for dichotomous manifest variates the trace-line or regression function takes the form (2.3.12)

$$p_i \mid x \equiv \mathcal{E}(y_i \mid x) = \sum_{\nu=0}^{r} a_{i\nu} x^\nu, \qquad j = 1, \cdots, n.$$

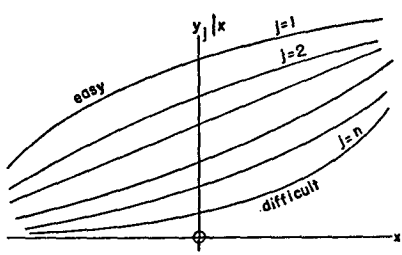


Figure 4.1.1

In introducing this model, Lazarsfeld (1950) did not develop solutions for it, except for the linear case where he has given formulae that are equivalent to those of a Spearman general-factor analysis. More generally, assuming a single latent variate, he noted that the rank of the covariance matrix (or *cross product* matrix, in his terms) would be equal to the highest degree of the trace-line polynomials. Lazarsfeld (1950, pp. 371–372) showed in particular for the quadratic case that

$$\mathcal{E}(y_i) \equiv p_i = a_{i0} + a_{i1}\mu_1 + a_{i2}\mu_2 \tag{4.2.1}$$

and

$$\begin{aligned}\operatorname{cov}(y_i, y_k) \equiv p_{ik} - p_i p_k &= a_{i1}a_{k1}(\mu_2 - \mu_1^2) \\ &+ a_{i2}a_{k2}(\mu_4 - \mu_2^2) + (a_{i1}a_{k2} + a_{k1}a_{i2})(\mu_3 - \mu_1\mu_2)\end{aligned} \tag{4.2.2}$$

where, as previously, $\mu_p = \mathcal{E}(x^p)$. He did not, however, go on to develop more general relations of this kind, or show how they might be used. In the ten years or so since this work of Lazarsfeld, there does not seem to have been any further contribution to the problem. One might guess that the main barrier to further work has been the occurrence of terms like the third on the right-hand side of (4.2.2). However, when we rewrite (2.3.12) in the form

$$p_i \mid x = \sum_{p=0}^{r} b_{ip}h_p(x),$$

where the $h_p(x)$ form an orthonormal set, the difficulty disappears. The model is then recognizable as the single-factor polynomial treated in sect. 3.2, with the dichotomous observations treated as measured data. An obvious advantage in using the methods of sect. 3.2 to fit the latent polynomial to empirical data is that these methods include a test of the model against the alternative multifactor linear model or latent hyperplane. This means that the assumption made by Lazarsfeld (1950) of a single latent variate in the latent polynomial model can be tested.

There are two possible embarrassments which may arise in the treatment of dichotomous observations in terms of the latent polynomial model. Firstly, it is possible that some of the fitted trace-lines may turn out to be nonmonotonic, within the interval in which the density of the latent distribution is not negligible, in cases where this is not credible on psychological grounds. Secondly, within this interval, one may obtain values for the conditional probability of a response to the item which lie outside the permissible range of zero to unity. One way to avoid these embarrassments is to employ models such as the normal ogive or the latent distance model, which are well behaved by definition in these respects. There is a good case for using a preliminary analysis in terms of the latent polynomial, however. Such a solution can then be converted if desired to a normal ogive or latent distance solution as shown in sect. 4.4 and 4.5.

4.3 *The Latent Class Models*

It was noted in the last section that Lazarsfeld (1950) in his early work on LSA introduced the latent polynomial model but did not develop its properties in any detail. Instead, he recommended replacing the polynomial trace-lines by an equivalent set of latent classes. The latent class model can be derived more generally (cf. sect. 2.3) from the most general form of the principles of LSA simply by ascribing a discrete distribution to the latent variate or set of latent variates.

Two forms of this model have been recognized. The latent class model proper is regarded as a model for dichotomous manifest variates. The treatment of this model represents the most advanced aspect of LSA (Green, 1951; Anderson, 1954; Gibson, 1955; Gibson, 1959; McHugh, 1956; Madansky, 1960; McDonald, 1962a). The case in which the manifest variates consist of multivalued measured data has been named the *latent profile* model by Gibson (1959). However, as argued in sect. 2.3, the distinction between the two models is not fundamental. Maximum-likelihood estimators (McHugh, 1956) have been supplied for the case of dichotomies, but not as yet in the more general case. For the present purposes, the two can be treated together as the *general latent class* model (McDonald, 1962a).

There are at least two important theoretical questions relating to the latent class model for which there has been no solution up to the present. The first is whether a "genuine" discrete distribution of the latent variates can be detected if it exists. (By current methods, it is always possible to *impose* such a distribution.) The second concerns the arrangement of the latent classes in the factor space. There are relations between the latent polynomial, the latent hyperplane, and the latent class model, such that in some ways they can be thought of as "equivalent" (Lazarsfeld, 1950). Thus, the latent polynomial of degree r, the linear model in r statistically independent latent variates, and the latent class model with $r + 1$ latent classes all entail a reduced covariance matrix of rank r. This is, of course, related to the fact that a minimum of $r + 1$ points will serve to define a polynomial of degree r in a single variable or to define an r-space. Existing methods of latent class analysis do not test whether the latent variates have a discrete distribution, or, in case they have, whether the points occupied by *latent classes* are distributed throughout r-space, or alternatively whether they span the r-space by lying on a curve within it, as assumed in effect by Lazarsfeld (1950) in treating the latent polynomial and latent class models as "equivalent." There is an early attempt by Lazarsfeld (1950, pp. 446–454) to deal with this problem. However, it rests on the somewhat restrictive assumption that the regression functions have at most a single maximum over the range of the distribution of the latent variate. The present approach supplies a partial answer only to the first of the above questions. The second is answered independently of the first, by the general theory of nonlinear factor analysis.

As the starting point for this aspect of the theory, we again consider the component variates v in (3.1.19) of the general theory. These are represented as the sum of a "true" part w and disturbances d. If the latent variate (or variates) x takes the values x_i with probabilities g_i $(i = 1, 2, \cdots)$, the corresponding "true" components w take values w_i with probabilities g_i. The disturbances d were assumed normally distributed, with a variance-covariance matrix given by (3.1.24). The component variates v will then have a density function which results from the combination of a discrete density function with a normal density function whose parameters can be estimated from a given sample. For a population in which the latent variates have a point distribution, we could expect to find in a sample from it that the distribution of the component variates is multimodal if the disturbances are small enough. A possible approach to the detection of a latent class case, then, is to look for such a multimodal distribution and to attempt to resolve the distribution into a discrete component due to the latent variates and a normal component due to the disturbances. To consider the feasibility and the limitations of such a procedure, we treat only the simplest case, involving two latent classes and dichotomous observations. This model is the *latent dichotomy* of Lazarsfeld (1950).

For the latent dichotomy, the following relations are easily obtained (see Appendix 4.3A). In this case the reduced covariance matrix of the manifest dichotomies is of rank unity, and the component variate v is a scalar. The "true" component w is identical with the latent variate x, and takes the two values $x_1 = -\sqrt{g_2/g_1}$ and $x_2 = \sqrt{g_1/g_2}$ with probabilities g_1 and g_2 respectively, with, of course, $g_1 + g_2 = 1$. Writing as usual $p_j \mid x_1$, $p_j \mid x_2$ for the probabilities of the positive response to item j given respectively that $x = x_1$, $x = x_2$, we introduce

$$\delta_j \equiv p_j \mid x_2 - p_j \mid x_1. \tag{4.3.1}$$

We then find that

$$\text{var } (d) = \frac{\sum_{j=1}^{n} \delta_j^2 \{a_j(1 - a_j) - g_1 g_2 \delta_j^2\}}{g_1 g_2 \left(\sum_{j=1}^{n} \delta_j^2\right)^2}, \tag{4.3.2}$$

where d, the disturbance component, is a scalar here. Given also the separation of the two values of w,

$$w_2 - w_1 = \sqrt{g_1/g_2} + \sqrt{g_2/g_1}, \tag{4.3.3}$$

by making further simplifying assumptions we can speculate quantitatively about the detectability of a latent dichotomy on the basis of a sample of data. This requires a determination of the amount of overlap to be expected between the two normal curves, with the given variance and separation

between their means. If we suppose the parameters δ_i to be equal, and $a_i = g_i$, for all items and set $\delta_i = \delta$ in (4.3.1), we obtain

$$\text{var } (d) = (1 - \delta^2)/n\,\delta^2. \tag{4.3.4}$$

In a practical case, g_1, g_2 would not be expected to vary outside the range 0.2 to 0.8. This would give a range for the separation $w_2 - w_1$ of 1.75 to 2.00 units, a variation so slight that it can reasonably be ignored as a first approximation. Taking the figure of 2.00 for the separation, we see for example with $n = 20$, and a probability-difference δ between the classes of 0.6, that the population *SD* for the disturbance component would be 0.669. Hence the resultant distribution of v would be strongly bimodal though with some overlap of the underlying normal curves as in Fig. 4.3.1 (see following example).

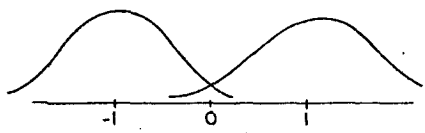


Figure 4.3.1

In principle one could introduce a cutting point on the v-dimension to divide a sample of observations on v into two groups, with a proportion of misclassified observations which is a function of var (d) and $w_2 - w_1$. If the within-group variation corresponds fairly closely to var (d), and the group means correspond reasonably to the expressions above for x_1, x_2, the latent class interpretation of the data will seem plausible. If there is any overlap at all, misclassification of observations will lead to a biasing downwards of the estimate of within-group variation. In principle, one could correct this bias.

It would appear then that the difference between a discrete latent class case and its equivalent continuous model can be detected, but the distinction may often remain uncertain. More generally, a strongly multimodal distribution of v, whether in one or more dimensions, with the modes disposed on a curve in the space or in a general configuration will lead us to suspect such a discrete distribution. In optimal cases where there are well-defined groups with minimal overlap, we would hope to fit the parameters of the latent class model by the present methods.

We would first fit the appropriate linear or nonlinear model

$$\mathcal{E}(y_i \mid x) = \varphi_i(x)$$

by the methods of Chap. 3. (These in no way depend on the continuity or or otherwise of x.) We would then estimate the density function g_i, $i = 1, \cdots, r$, (the proportion of subjects in each latent class) by simple enumeration of the number of subjects in each group in the sample distribution of v.

The remaining parameters required are the $n \times r$ quantities $\varphi_i(x_i)$. These are obtained by substituting group means for x in $\varphi_i(x)$. In the case of manifest dichotomies, each $\varphi_i(x_i)$ is interpreted, as usual, as the conditional probability of endorsing item j for a "subject" belonging to latent class i. For multivalued manifest variates each $\varphi_i(x_i)$ represents a latent class mean as in Gibson's (1959) discussion of the "latent profile model." Since the parameters g_i and $\varphi_i(x_i)$ of the latent class model are independent of the arrangement of the latent classes in the factor space, the problem of rotation does not arise. However, if we proceed in this way, information as to the configuration of the classes has necessarily been obtained and will form a valuable supplement to the latent class description.

In fact, this procedure is the converse of Gibson's. Gibson (1959), as noted in sect. 2.5, proposes that we first obtain a latent class solution, and then arrange the classes in space in some way, to obtain a nonlinear factor model. The second step is necessarily quite arbitrary. In contrast, here we first obtain a nonlinear factor model that fits the data, and then *if groupings* are observable in the arrangement of the component variates in the factor space, we permit ourselves a latent class interpretation. Both methods can fail, in the sense of imposing an incorrect model without being able to know that it is incorrect. In general, however, it would seem fairly harmless to impose a continuous model on data which is "really" discrete in its latent structure, but perhaps not quite so harmless to impose a latent class model on data which is "really" continuous. Further, the second procedure is less informative.

For the purposes of the following artificial example, it is useful to have expressions for the parameters of the linear model which correspond to a given latent dichotomy. It is shown in Appendix 4.3A that if, in the linear model for manifest dichotomies,

$$\varphi_i(x) \equiv \mathcal{E}(y_i \mid x) = a_i + b_i x,$$

x has a point distribution, taking only the values x_1 and x_2 with probabilities g_1, g_2 respectively, then

$$a_i = \varphi_{i1} g_1 + \varphi j_2 g_2 \tag{4.3.5}$$

and

$$b_i = \sqrt{g_1 g_2}(\varphi_{i2} - \varphi_{i1}) = \sqrt{g_1 g_2}\ \delta_i, \tag{4.3.6}$$

where $\varphi_{i1} \equiv \varphi_i(x_1) = p_i \mid x_1$, $\varphi_{i2} = \varphi_i(x_2) = p_i \mid x_2$, and g_1 (or g_2) are the three parameters of the latent dichotomy.

Numerical Example 4.3a

As an illustration of the preceding discussion, a set of artificial data was constructed, using a table of random numbers, to correspond to a latent dichotomy.

The intention was to construct an optimal case, in which one could expect to find good separation of the groups, in terms of the procedures outlined above. This leads to a case that looks rather trivial at first sight, but it serves to bring out a point not mentioned above, namely, the relative advantages of using the component variate v over the use of total score on the n items. Clearly, in data conforming to a latent dichotomy one might expect that the score totals for each subject over the n items in the set would tend, like the component variate v, to yield a bimodal distribution. It might be supposed that no more elaborate machinery than this is required to detect a latent dichotomy when it exists. As will appear, the component variate distribution has some advantages over the total score for this purpose.

A score matrix was constructed consisting of one hundred "subjects" by twenty variables ("items"). Each variable was scored unity if a pair of digits in the table of random numbers was greater than a criterion number, and zero otherwise. A sample was constructed with $g_1 = g_2 = 0.5$ for all items. For items one through ten the criterion number eighteen was taken for fifty of the "subjects," with eighty for the other fifty. This is approxi-

TABLE 4.3.1

Parameter Estimates: Latent Dichotomy

Item	$p_j = a_j$	$f_j = b_j$	Estimates using group means		Estimates using group proportions	
			φ_{j1}	φ_{j2}	φ_{j1}	φ_{j2}
1	0.49	0.403	0.095	0.887	0.087	0.893
2	0.49	0.346	0.151	0.831	0.144	0.836
3	0.52	0.317	0.210	0.832	0.203	0.837
4	0.46	0.304	0.162	0.759	0.156	0.764
5	0.48	0.288	0.198	0.764	0.192	0.768
6	0.50	0.298	0.208	0.794	0.202	0.798
7	0.49	0.283	0.213	0.769	0.207	0.773
8	0.44	0.390	0.058	0.824	0.050	0.830
9	0.48	0.378	0.110	0.852	0.102	0.858
10	0.52	0.308	0.218	0.823	0.212	0.828
11	0.43	−0.036	0.465	0.395	0.466	0.394
12	0.43	−0.054	0.485	0.377	0.484	0.376
13	0.52	−0.020	0.540	0.500	0.540	0.500
14	0.47	−0.098	0.566	0.373	0.568	0.372
15	0.61	0.047	0.564	0.656	0.563	0.657
16	0.50	0.048	0.453	0.547	0.452	0.548
17	0.47	−0.037	0.506	0.434	0.507	0.433
18	0.54	−0.028	0.568	0.512	0.568	0.512
19	0.48	0.023	0.458	0.503	0.457	0.503
20	0.53	−0.067	0.596	0.464	0.597	0.463
		Mean (1–10)	0.162	0.813	0.155	0.818
		Mean (11–20)	0.520	0.476	0.520	0.476

TABLE 4.3.2

Distributions of "Subjects": Latent Dichotomy

Distribution of v		Distribution of score totals			
C.I.	f	Score	f	f [Latent class 1]	f [Latent class 2]
1.80 – 1.99		20			
1.60 – 1.79		19	1		1
1.40 – 1.59	7	18	1		1
1.20 – 1.39	9	17	1		1
1.00 – 1.19	13	16	4		4
.80 – .99	5	15	4		4
.60 – .79	8	14	11		11
.40 – .59	5	13	9		9
.20 – .39	3	12	11		11
.00 – .19		11	4		4
−.20 – −.01		10	5	2	3
−.40 – −.21	5	9	7	6	1
−.60 – −.41	6	8	11	11	
−.80 – −.61	7	7	9	9	
−1.00 – −.81	8	6	10	10	
−1.20 – −1.01	9	5	5	5	
−1.40 – −1.21	7	4	6	6	
−1.60 – −1.41	6	3	1	1	
−1.80 – −1.61	2	2			
−2.00 – −1.81		1			

mately the optimal case of a latent dichotomy for which there is a possible equivalent latent linear model, with a rectangular latent distribution, in terms of (4.3.5) and (4.3.6). If we take any greater difference than this between φ_{j1} and φ_{j2}, the equivalent latent linear model would give values of less than zero and greater than one for the probability of a response, within the range of x in which the population density is nonzero. For items eleven through twenty, the criterion number was forty-nine for all subjects, corresponding to a conditional probability of 0.5 for either response in both latent classes. Thus the total set of items consists of ten "discriminating" and ten "nondiscriminating" items as between the two latent classes.

The covariance matrix is given in Appendix 4.3B. This is satisfactorily accounted for by a general factor, with loadings as given by Table 4.3.1. The distribution of the component variate is given in Table 4.3.2. The distribution falls into two clearly separated groups.

Pretending ignorance of the origins of this set of data, we can draw the following conclusions.

(1) Supposing that there is no misclassification, since there are fifty subjects per group, the best estimate of the proportions in the two presumptive latent classes will be $g_1 = g_2 = 0.5$.

(2) The proportion p_i of positive responses, given in Table 4.3.1 and the factor loadings there given, provide the parameters a_i and b_i of the latent linear model. Substituting the group means, 0.985 for the upper group and -0.979 for the lower group, as the two values of x in

$$\varphi_i(x) = a_i + b_i x$$

gives the estimates of φ_{i1} and φ_{i2} in Table 4.3.1. In the absence of sampling theory, the fit to the theoretical values $\varphi_{i1} = 0.19$ and $\varphi_{i2} = 0.81$ for items one through ten and $\varphi_{i1} = \varphi_{i2} = 0.50$ for items eleven through twenty does not seem too bad. Separate means of these parameters are also given in Table 4.3.1 for items one through ten and items eleven through twenty.

(3) Instead of using the two values of x estimated from the group means, we could take the values $x_2 = \sqrt{g_2/g_1} = 1$, $x_1 = -\sqrt{g_1/g_2} = -1$, using the estimated proportions g_1, g_2 above, to estimate φ_{i1}, φ_{i2}. These alternative estimates are given also in Table 4.3.1.

(4) We also note that the variances within the upper and lower presumptive groups are, respectively, 0.109 and 0.153. These may be compared with the theoretical population variance for the disturbance component, viz. 0.151, given by substituting $\delta_i = 0.81 - 0.19$ (items one through ten) and $\delta_i = 0$ (items eleven through twenty) in (4.3.2). They may also be compared with the further estimate 0.095, obtained by subtracting unity from the total variance of the component variate.

From this much it would seem that at least in an optimal case the procedures employed as a matter of routine in nonlinear factor analysis can reveal a genuine latent class case if it exists, and can be made to give a reasonable fit to the parameters.

In this example, a simpler procedure would have sufficed to give us grounds for suspecting a latent dichotomy. Table 4.3.2 gives the distribution of the total scores over the twenty items for the one hundred "subjects" (and also the separate distributions for the two latent classes, as known on the basis of the previous work). This distribution is in fact suspiciously bimodal. There is, however, considerable overlap in the distributions of the score totals.

More generally, the weight given to an item in determining the value of the component variate v is proportional to its δ_i value, in the present notation. Hence the peaks of the distribution of v obtained in a latent dichotomy case will be sharpened in comparison with the distribution of score totals, so that the distribution of v could enable one to detect a latent dichotomy when there is little or no suggestion of this in the distribution of the score totals.

4.4 *Normal-Ogive Model*

The normal-ogive model is a basic and important model in mental test theory. Both the early history of the notions involved in the model, dating at least from Richardson (1936) and Guilford (1936), and the elegant formal treatment of it by Lord (1953) lie outside the framework of latent structure analysis. Nevertheless, the model can be described as a particular case in LSA, as indicated in (2.3.14).

It is not necessary to review the existing developments of this model in any detail here. It may suffice to note that Lord (1953) has deduced the maximum-likelihood estimators for the parameters of the model; also, among other interesting properties, it has been shown that if the latent variate is normally distributed, the model should yield a matrix of tetrachoric correlations which is of rank unity.

The quite limited aim of this section is to show that given an initial treatment of a set of observations in terms of the single-factor polynomial, it is possible to transfer from such a solution to the parameters of the normal-ogive model by way of a Fourier analysis of the normal ogive. The major advantage of this approach is of course that the transfer can be carried out as an afterthought, having first established that the set of items depends on only a single latent variate, and that the shapes of the trace functions are reasonably compatible with the normal-ogive model. A good deal of further work will be required, however, before it is known whether the estimates of the model parameters so obtained are at all well behaved.

We consider, then, the regression function as in (2.3.14),

$$\varphi_i(x) \equiv \mathcal{E}(y_i \mid x) = N\{(x - \mu_i)/\sigma_i\},$$

where

$$N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty} \exp\left(\frac{-z^2}{2}\right) dz.$$

While the argument would hold for quantitative manifest variates, one may as well think of the observations as items, in this case, and take the regression functions to represent the conditional probability of endorsing the item for a given position on the latent dimension.

As indicated in sect. 3.7, the formal Fourier expansions (3.7.1) of the $\varphi_i(x)$ are

$$\varphi_i(x) \sim c_{i0}h_0(x) + c_{i1}h_1(x) + \cdots + c_{ip}h_p(x) \cdots$$

where the Fourier coefficients (3.7.2) are given by

$$c_{ip} = \int \varphi_i(x)h_p(x)g(x)\, dx$$

with the density function of x given by $g(x)$. These have the least-squares property (3.7.3). For any given $g(x)$, with the above specification of $\varphi_i(x)$, the c_{ip} can be written as functions in μ_i and σ_i. A common assumption is that x is normally distributed. Under this assumption it can be shown (see Appendix 4.4A) that

$$
\begin{aligned}
(4.4.1) \qquad c_{i0} &= N\{-\mu_i/(1+\sigma_i^2)^{1/2}\} \\
c_{ip} &= p^{-1/2}(1+\sigma_i^2)^{-p/2} h_{p-1}\{\mu_i/(1+\sigma_i^2)^{1/2}\} n\{\mu_i/(1+\sigma_i)^{1/2}\} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad p = 1, 2, \cdots,
\end{aligned}
$$

where

$$
n(t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right)
$$

and the $h_p(\cdot)$ are obtained by normalizing the Hermite–Tchebycheff polynomials. The first four coefficients, which are probably all that would ever be required in practice, are given by

$$
\begin{aligned}
(4.4.2) \qquad c_{i0} &= N\{-\mu_i/(1+\sigma_i^2)^{1/2} \\
c_{i1} &= (1+\sigma_i^2)^{-1/2} n\{\mu_i/(1+\sigma_i^2)^{1/2}\} \\
c_{i2} &= \frac{1}{\sqrt{2}}(1+\sigma_i^2)^{-1} \frac{\mu_i}{(1+\sigma_i^2)^{1/2}} n\{\mu_i/(1+\sigma_i^2)^{1/2}\} \\
c_{i3} &= \frac{1}{\sqrt{3}}(1+\sigma_i^2)^{-3/2} h_2\{\mu_i/(1+\sigma_i^2)^{1/2} n\{\mu_i/(1+\sigma_i^2)^{1/2}\}.
\end{aligned}
$$

Let us suppose, then, that we have data which conform to the normal-ogive model. We fit the single-factor polynomial to it by the previous methods, and obtain thereby

$$
(4.4.3) \qquad p_i \mid x \equiv \mathcal{E}(y_i \mid x) = \sum_{p=0}^{r} b_{ip} h_p(x).
$$

If in addition we find that the distribution of x can be considered normal, so that the polynomials $h_p(x)$ approximate the normalized Hermite–Tchebycheff series, under these conditions it would seem reasonable to make the identification

$$
c_{ip} = b_{ip}.
$$

Then by (4.4.1) we have

$$
\begin{aligned}
(4.4.4) \qquad \mu_i/(1+\sigma_i^2)^{1/2} &= N^{-1}(b_{i0}) \\
(1+\sigma_i^2)^{1/2} &= \frac{1}{b_{i1}} n\{N^{-1}(b_{i0})\}
\end{aligned}
$$

whereby the parameters μ_i and σ_i can be computed given the values of b_{i0} and b_{i1}. If the solutions (4.4.3) have yielded higher order terms such as b_{i2}, b_{i3}, b_{i4}, the compatibility of these with (4.4.1) can be checked. Alternatively, estimates using these in (4.4.1) might be combined with those from (4.4.4) to improve estimation in practice.

A good deal of further work will be necessary before these proposals can be considered workable. For the present, the following points will serve to conclude this section. Firstly, if the latent linear model gives a reasonable fit to the data, it is still possible to transfer to the normal-ogive model by way of (4.4.4). Nonlinearity is not necessary. Secondly, if the assumption that the latent variate is normally distributed breaks down, it should be possible to develop analogues to (4.4.1) post facto, according to the distribution of x which is indicated by the data. In some cases, numerical integration may be necessary, and the problem of inverting the relations as in (4.4.4) may be intractable.

One feature of the normal-ogive model could be considered a weakness in it. The upper and lower asymptotes of the conditional probability functions $\varphi_i(x)$ for the observed dichotomies are respectively unity and zero. Thus, in an ability context, no allowance is made for success due to guessing by a subject of low ability, or of failure due to "chance" errors in a subject of high ability. It would be convenient for some applications to relax these restrictions on the model, by introducing upper and lower asymptotes which can be determined from the data, and which are free to take values other than zero and unity. It is very easy to extend the present treatment of the model to cover this point. It is shown in Appendix 4.4A that if we introduce the parameters λ_{i0} for the lower limit and λ_{i1} for the upper limit of the ogive function, we simply replace (4.4.1) by

$$
\begin{aligned}
(4.4.5) \qquad c_{i0} &= \lambda_{i0} + (\lambda_{i1} - \lambda_{i0})N\{-\mu_i/(1 + \sigma_i^2)^{1/2}\} \\
c_{ip} &= (\lambda_{i1} - \lambda_{i0})p^{-1/2}(1 + \sigma_i^2)^{-1/2}h_{p-1}\{\mu_i/(1 + \sigma_i^2)^{1/2}\} \\
&\qquad \cdot n\{\mu_i/(1 + \sigma_i^2)^{1/2}\}, \qquad p = 1, 2, \cdots .
\end{aligned}
$$

Given estimates of the parameters c_{ip} up to the third order from a nonlinear factor analysis, it is possible as previously to estimate the parameters of this model, μ_i, σ_i, λ_{i0} and λ_{i1} from these, since

$$
(4.4.6) \qquad \frac{c_{i2}}{c_{i1}} = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{1 + \sigma_i^2}},
$$

$$
(4.4.7) \qquad \frac{c_{i3}}{c_{i1}} = \sqrt{\frac{2}{3}} \cdot \frac{1}{\sqrt{1 + \sigma_i^2}} \frac{h_2\{\mu_i/\sqrt{1 + \sigma_i^2}\}}{\mu_i/\sqrt{1 + \sigma_i^2}}.
$$

Given the values of μ_i, σ_i from these relations, the remaining parameters λ_{i0} and λ_{i1} are readily determined from the equations for c_{i0} and c_{i1}.

Since the solution for this version of the normal-ogive model requires terms in the initial nonlinear factor analysis up to the cubic, a good deal of empirical work will probably be required before it is known whether these relations can generally be expected to yield sensible results.

4.5 *Latent Distance Model*

In the latent distance model (2.3.15) we have

$$\varphi_i(x) \equiv \mathcal{E}(y_i \mid x) = \gamma_i, \qquad x \leq x_i$$
$$= \beta_i, \qquad x > x_i,$$

i.e., a "step" or "jump" function for the trace-line or regression of item j on the single latent variate x. This model was developed by Lazarsfeld (1950) as one possible probability-analogue of the perfect scale. Methods for determining the parameters of the model have been described by Lazarsfeld (1950) and Hays and Borgatta (1954). It should be noted that these methods require a preliminary ordering of the items in respect of their *breaking points* x_i.

As in the case of the normal-ogive model just treated, the object here is to give the results of a harmonic analysis of the step-function in terms of orthonormal polynomials. This, again, yields a set of relations by means of which an initial solution in the form of a latent polynomial can be made to yield the parameters of the latent distance model. One advantage in this case is that the preliminary ordering of the items in respect of x_i is not necessary in the indirect method.

With the notation of the previous section, and with parallel remarks, for this case we obtain (see Appendix 4.5A),

$$c_{i0} = \beta_i + (\gamma_i - \beta_i)N(x_i) \tag{4.5.1}$$

$$c_{ip} = \frac{1}{\sqrt{p}}(\beta_i - \gamma_i)h_{p-1}(x_i)n(x_i), \qquad p \geq 1,$$

where the h_p are the normalized Hermite–Tchebycheff polynomials, under the assumption that the latent variate has a normal density function. The first few coefficients c_{ip} are

$$c_{i1} = (\beta_i - \gamma_i)n(x_i) \tag{4.5.2}$$

$$c_{i2} = \frac{1}{\sqrt{2}}(\beta_i - \gamma_i)x_i n(x_i)$$

$$c_{i3} = \frac{1}{\sqrt{3}}(\beta_i - \gamma_i)\frac{x_i^2 - 1}{\sqrt{2}}n(x_i).$$

Again by identifying the coefficients b_{ip} in the latent polynomial analysis with the Fourier coefficients c_{ip} we can in principle solve for the parameters

of the latent distance model. The first three terms are sufficient, since by (4.5.2) we can estimate x_i by

$$x_i = \sqrt{2}\,\frac{b_{i2}}{b_{i1}}, \tag{4.5.3}$$

following which the values of β_i, γ_i can easily be obtained from the equations for $c_{i0}(= b_{i0})$ and $c_{i1}(= b_{i1})$. If the initial analysis has given higher-order terms, the parameters so obtained can be checked for consistency with these.

It may be noted that a restricted form of the latent distance model has been described (cf. Torgerson (1958)) in which it is assumed that

$$\gamma_i = 1 - \beta_i. \tag{4.5.4}$$

With this assumption, c_{i0} and c_{i1} are given by

$$\begin{aligned} c_{i0} &= \beta_i + (1 - 2\beta_i)N(x_i) \\ c_{i1} &= (2\beta_i - 1)n(x_i). \end{aligned} \tag{4.5.5}$$

Given estimates of the c_{i0}, c_{i1} from a linear factor analysis, (4.5.5) could be solved by a graphical or successive-approximation method to yield the parameters of the restricted latent distance model. The unrestricted version, on the other hand, requires an initial factor solution which is nonlinear, with terms at least up to the second degree, if it is to be treated by the present method.

As in the case of the normal-ogive model, if the assumption of a normal density function for the latent variate breaks down, it may be possible to find an appropriate expression for the observed distribution and carry out a corresponding Fourier analysis.

Again it should be pointed out that this procedure may require the use of numerical integration, and could prove intractable in some cases.

4.6 *The Perfect Scale*

As mentioned earlier, the *ideal answer pattern* first recognized by Walker (1931), and implicit in Ferguson's (1941) work on difficulty factors, has more recently been extensively studied by Guttman. A large body of literature exists, usually referring to it as the *perfect scale* or *Guttman scale*. The most important developments are contained in Loevinger (1947; 1957), Guttman (1950; 1954c), Lord (1958), Goodman (1959), Maxwell (1959a) and Schuessler (1961).

In keeping with the general intentions of the present approach, the perfect scale will here be considered only as a common limiting case of the two models just considered. As mentioned in sect. 2.3, in the normal-ogive model (2.3.14),

$$\varphi_i(x) = N\{(x - \mu_i)/\sigma_i\},$$

it is clear that as $\sigma_i \to 0$, in the limit we have

$$\varphi_i(x) = \begin{cases} 1, & x > \mu_i \\ 0, & x < \mu_i. \end{cases} \tag{4.6.1}$$

Similarly, in the latent-distance model (2.3.15),

$$\varphi_i(x) = \begin{cases} \beta_i, & x > x_i \\ \gamma_i, & x < x_i, \end{cases}$$

if we set $\gamma_i = 0$, $\beta_i = 1$ for all j, then

$$\varphi_i(x) = \begin{cases} 1, & x > x_i \\ 0, & x < x_i. \end{cases} \tag{4.6.2}$$

The parameters x_i, μ_i take on the same meaning in (4.6.1) and (4.6.2).

Provided that the values of x_i for the items are all distinct, the model (4.6.2) gives rise to the ideal answer pattern or perfect scale. It is worth noting that, as is to be expected, the Fourier coefficients given by (4.4.1) and (4.5.1) reduce in this case to identical expressions, namely,

$$\begin{aligned} c_{i0} &= N(-x_i) \\ c_{ip} &= p^{-1/2} h_{p-1}(x_i) n(x_i), \qquad p \geq 1. \end{aligned} \tag{4.6.3}$$

These correspond to the principal component weights in the more general theory given by Guttman (1950), and the normalized Hermite–Tchebycheff polynomials $h_p(x)$ correspond to the principal component scores, for the case where the latent variate has a normal distribution.

In contrast to the psychological interpretation of the principal components given by Guttman (1950; 1954c) and Suchman (1950), the components are here treated simply as the harmonics of a step function. In terms of the present theory, if we were to fit a latent polynomial to a set of scalable data, this would yield a linear combination of the Guttman principal components which gives the best approximation in the sense of least squares to a step-function.

4.7 *The Simplex Property and Translated Function Models*

In this section we examine certain limited aspects of the simplex theory developed by Guttman (1954b; 1955b). As mentioned in sect. 2.3, a correlation matrix has the *additive simplex* property, if (2.3.22),

$$r_{jk} = \alpha_j / \alpha_k, \qquad j \leq k$$

for a specified ordering of the variables. A more general definition of the simplex property (Guttman, 1955b), which allows negative correlations to occur, is given by

$$\mathcal{E}\{(y_j - y_k)(y_k - y_l)\} = 0, \qquad j \leq k \leq l, \tag{4.7.1}$$

i.e., the sequence of random variables $y_1, y_2, \cdots, y_n$ has *orthogonal increments*. We consider the additive simplex only.

Among the mathematical properties of a simplex matrix that have been noted by Guttman, there are two of particular importance. When the variables are appropriately ordered, any second-order minor vanishes if its elements are all drawn from one side of the leading diagonal, but does not vanish in general if elements are drawn from both sides. Also, the largest correlations are adjacent to the leading diagonal, and their values decrease progressively as one moves away from the diagonal. One difficulty for empirical applications of the simplex model is that it is easy to find other models which entail the second of these properties and closely approximate the first, yet do not possess the properties (2.3.22) or (4.7.1).

It was pointed out earlier that the correlation matrix (phi-coefficients) of the perfect scale possesses the simplex property. In this case the principal components of the scale can be regarded as the harmonics of a step-function as in the previous section. While the simplex property does not entail a unique factor model giving rise to it, Guttman (1954b, p. 323) makes the additional assumption in his simplex theory that the observed variates are continuously measurable, with a joint normal density function and hence linear mutual regressions. Because of this additional assumption, he argues that the principal component scores (equivalent to the component variates v of Chap. 3) must be mutually statistically independent. This is in contrast to the principal component scores for the perfect scale, for which case he finds relations between them similar to those between the $h_p(x)$ in Chap. 3. It appears that neither Guttman's additional assumption nor its further implications have yet been tested on any empirical simplex matrix.

Given the assumption of a normal multivariate distribution for the observed variates, the most plausible factor model giving rise to the simplex property (4.7.1) is, in the present notation,

$$y_i = \sum_{p=1}^{i} a_p x_p, \tag{4.7.2}$$

that is,

$$y = xA, \tag{4.7.3}$$

where

$$A \equiv \begin{bmatrix} a_1 a_1 a_1 & \cdots & a_1 \\ & a_2 a_2 & \cdots & a_2 \\ & & a_3 & \cdots & a_3 \\ & & & \ddots & \\ & & & & a_n \end{bmatrix}$$

is an $n \times n$ matrix with zero elements everywhere below the leading diagonal (cf. Guttman, 1954b, pp. 309–311) and nonzero elements, repeated as shown, elsewhere, and $x_1, \cdots, x_n$ are independent normal, with mean zero, variance unity. Since A is in triangular form, it is nonsingular, hence the correlation matrix is of full rank. It is in terms of this particular factor model for the simplex that Guttman interprets the order of the observed variates as a progressive increase in "complexity." In the field of abilities, each successive test is thought of as requiring all the abilities called upon in the preceding test, plus one more.

In Guttman's theory, a perfect simplex is one in which the simplex property holds precisely. In a *quasi-simplex,* an error term is added, and it becomes necessary to adjust the elements in the leading diagonal of the correlation matrix as in preceding factor theory. In general, such adjustments *as required by* the error-term in the factor model should not reduce the rank of the correlation matrix if the simplex condition holds.†

Guttman (1957b) lists twenty-six sets of variables which appear to have a quasi-simplex form. These were discovered as submatrices of published empirical correlation matrices.

Humphreys (1960) draws attention to the existence of the quasi-simplex property in data obtained by retesting subjects at intervals over time in certain maturation or learning studies. Here, the model (4.7.2) seems applicable. The successive factors x_p can be thought of as increments in growth or learning which are statistically independent of the existing level of development. Such a situation resembles certain stochastic processes of the random-walk type (cf. Feller, 1957, pp. 65–85).

In contrast to the evidence as interpreted by Guttman, it has been noted by Borgatta (1958) and Humphreys (1960) that with a minimum of two factors it is possible to generate a matrix which closely resembles a simplex, in that the two conditions mentioned above are approximately satisfied. Following Humphreys, we will refer to such a matrix as a pseudo-simplex. There is a fundamental distinction in theory between a pseudo-simplex and a quasi-simplex, but in typical empirical work it may prove impossible to distinguish between them at all.

An equally important distinction, especially in the case of learning and

† It has been shown very ingeniously by DuBois (1960) that a correlation matrix having the simplex property (without error) can be reduced in rank to half its order by suitable adjustments to the diagonal elements. This demonstration does not seem to have any fundamental significance, however. In the usual situation with a factor model plus error, we describe the n observed variates in terms of r common factors plus n "specific factors" ascribable to unique "errors." In contrast, DuBois is able to show that an error-free simplex matrix can be described in terms of $n/2$ common factors and $n/2$ specific factors. That is, there is a rotation of the triangular matrix A within the n-space such that half of its column-vectors each have only one nonzero element. While the issue may be worth exploring further, it does not seem sensible to use the communality concept except on the basis of a model allowing for random errors.

maturation studies, is that between the multifactor simplex, with the factor model (4.7.2), and the unifactor simplex as in sect. 4.6. It is shown in Appendix 4.7A that the factor model

$$y_i = f_i(x), \tag{4.7.4}$$

where

$$f_i(x) = \begin{cases} 0, & x < x_{i0} \\ \varphi_i(x), & x_{i0} \leq x \leq x_{i1} \\ 1, & x > x_{i1} \end{cases}$$

has a quasi-simplex property, in the sense that it possesses an approximation to the property (2.3.22), provided that, for any two successive variables j and k, $x_{k0} > x_{j1}$. This model satisfies (2.3.22) precisely, if and only if $\varphi_i(x)$ is a step-function, as in the argument of sect. 4.6. Model (4.7.4) corresponds to the situation where the observed variates have identical upper and lower limits, and where in terms of the latent variate x (the "ability," say), any observed variate leaves its "floor" after the preceding one reaches its ceiling.

A model such as (4.7.4) may seem unlikely to apply in real life. Yet some approximation to it may well hold in such cases as the learning and maturation data considered by Humphreys (1960), thus providing an alternative to the random-increments model. It may also hold approximately for subtests formed from sets of items of widely differing difficulty, as in the first empirical example in sect. 5.1 below, and in the analysis by Gabriel (1954) of the subtests of the 1937 Raven's Progressive Matrices Test.

An important application of nonlinear factor analysis would be to the task of distinguishing between unifactor and multifactor models for data which appear to have the simplex property. This would seem to be of particular importance in learning and maturation studies, where the theoretical distinction is between the random-increments model and one in which the simplex property is due merely to the scale properties of the manifest variates.

Translated Function Models

The step-function model (4.6.2) for the perfect scale can be described as

$$\varphi_i(x) = \varphi(x - x_i), \tag{4.7.5}$$

where $\varphi(x)$ is defined by

$$\varphi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \tag{4.7.6}$$

This immediately suggests a whole class of models defined by (4.7.5) obtainable by introducing different functions $\varphi(x)$ in place of (4.7.6). These may rea-

TABLE 4.7.1

1.000	.955	.449	−.318	−.600
.955	1.000	.707	.000	−.318
.449	.707	1.000	.707	.449
−.318	.000	.707	1.000	.955
−.600	−.318	.449	.955	1.000

sonably be called *translated function models,* as they are obtained by the translation of a prescribed function along the x-axis. The particular case of the normal-ogive model in which the item parameters σ_i (cf. sect. 4.4) are equal for all items is an important case of this class, in which the items differ in difficulty but not in discrimination.

In Appendix 4.7B some properties are given for the model in which a single-factor quadratic regression function is translated along the x-axis. With one particular choice of numerical values for the parameters, this model yields the correlation matrix shown in Table 4.7.1. This is a pseudo-simplex of rank two, corresponding to the Guttman (1955b) generalized simplex model. Closer approximations to a quasi-simplex could be obtained with models of this type containing terms of higher degree. It seems likely that the approach to translated function models in Appendix 4.7B is a theoretical blind alley. The important point about these models is that they serve to shed more light on Guttman's radex theory, considered as a theory of conditional order. Guttman has emphasized the fact that the simplex property, when it is found, yields a unique ordering of items or tests, and the significance of this ordering is independent of the factor model which underlies the data. In view of the above, it would seem that one cannot ignore the fundamental difference between an ordering of tests or items which rests on the translation of a regression function along a single axis, and an ordering which rests on the successive addition of independent factors. Nonlinear factor analysis seems to be the essential device for distinguishing between these alternatives.

4.8 *Curve-Fitting by Nonlinear Factor Analysis*

A further application of nonlinear factor analysis consists of an extension of one aspect of work by Tucker (1958; 1960) on the simultaneous fitting of a set of individual curves by linear factor analysis. In this section, an account of this extension is given, together with an alternative treatment based on classical methods.

We consider a matrix of observations

$$Y \equiv \langle y_{it} \rangle \quad \begin{matrix} i = 1, \cdots, n \\ t = 1, \cdots, m. \end{matrix}$$

This may be thought of as containing measures on each of n individuals under each of m levels of a prescribed quantitative treatment x_t. Other interpretations are possible. (To fix ideas, we may think of the matrix as

representing responses of individuals (i) after a number of learning trials (t)). We regard Y as a random sample of one realization from an ensemble of possible realizations of the process being studied with $x_1, \cdots, x_m$ fixed and known. This corresponds to Model II as discussed in sect. 2.1.

We are interested in a set of individual laws of the form

$$y_{it} = f_i(x_t) + e_{it} \tag{4.8.1}$$

where each e_{it} is a random deviate.

Tucker (1958; 1960) has proposed to describe any set of individual curves as linear combinations of a minimal set of *reference curves*, i.e., to write

$$y_{it} = \sum_{r=1}^{s} a_{ir} f_r(x_t) + e_{it}, \tag{4.8.2}$$

where the $f_r(x_t)$ are $s \leq m$ prescribed linearly independent functions in x_t and where, of course, $|f_r(x_t)| < \infty$ for all r, t. The coefficients a_{ir} are to be estimated from the sample. Tucker's approach to the problem is rather more flexible in some respects than the development of it to be given here, since a variety of functions might be prescribed, perhaps on theoretical grounds according to the given data, whereas here we will as previously consider only the treatment in terms of orthonormal polynomials. Cliff (1962) has examined the rotation problem that arises in Tucker's work and has provided a solution to it for cases where one has a criterion towards which the factor solution is to be rotated. The present approach to this problem, on the other hand, is a natural development of the methods of Chap. 3.

Method I: Procedure Using Nonlinear Factor Analysis

We may rewrite (4.8.2) in matrix form,

$$Y = AF(x) + E, \tag{4.8.3}$$

where

$$A \equiv \langle a_{ir} \rangle \quad \text{an } n \times s \text{ matrix},$$
$$F(x) \equiv \langle f_r(x_t) \rangle \quad \text{an } s \times m \text{ matrix},$$

of rank $s \leq m$, and

$$E \equiv \langle e_{it} \rangle, \quad \text{an } n \times m \text{ matrix}.$$

We assume* that

$$\begin{cases} \mathcal{E}(e_{it}) = 0 & \text{for all } i, t, \\ \mathcal{E}(e_{it}^2) = \sigma_i^2 & \text{for all } t, \\ \text{and} \\ \mathcal{E}(e_{it}e_{ju}) = 0 & \text{for all } i \neq j \quad \text{and} \quad \text{all } t, u. \end{cases} \tag{4.8.4}$$

* It may be noted that serial correlation between "trials" within "individuals" is allowed in the model, since in general

$$\mathcal{E}(e_{it}e_{iu}) \neq 0.$$

By (4.8.3) and (4.8.4) we have

$$\mathcal{E}(YY') = AF(x)F'(x)A' + U^2, \tag{4.8.5}$$

where

$$U^2 = m \operatorname{diag} \{\sigma_1^2 \cdots \sigma_n^2\}.$$

In general,

$$F(x)F'(x) \neq I,$$

so that one might consider it necessary to employ an oblique factor solution. However, as in the corresponding equations (3.1.5) through (3.1.10), since $F(x)$ is of full rank, there exists a nonsingular matrix T such that

$$TF(x)F'(x)T' = I. \tag{4.8.6}$$

Writing

$$G(x) = TF(x), \tag{4.8.7}$$

whence

$$F(x) = T^{-1}G(x), \tag{4.8.8}$$

we have from (4.8.3)

$$Y = AT^{-1}G(x) + E \tag{4.8.9}$$

and writing

$$B = AT^{-1} \tag{4.8.10}$$

we obtain

$$Y = BG(x) + E, \tag{4.8.11}$$

where

$$G(x)G'(x) = I. \tag{4.8.12}$$

Explicitly, (4.8.11) may be written

$$y_{it} = \sum_{r=1}^{s} b_{ir}g_r(x_t) + e_{it}. \tag{4.8.13}$$

Hence without loss of generality the representation (4.8.2) can be reduced to the representation (4.8.13), where the functions $g_r(\cdot)$ are orthonormal.

By (4.8.4), (4.8.11) and (4.8.12),

$$\mathcal{E}(YY') = BB' + U^2, \tag{4.8.14}$$

where B is of full rank.* Also, (4.8.11) has the same form as (3.1.10). The important difference is that in the present case the x_t are *known*. Since in this case the $g_r(x_t)$ are functions in a single variable, we have in particular the single-factor case, hence in principle it can be subjected to a first analysis by the methods of sect. 3.2. Such an analysis will yield functional relations between the g_r of the type

$$g_r(x) = h_r\{g_1(x)\} \tag{4.8.15}$$

where the functions $h_r(\cdot)$ are the polynomials of (3.2.1).

In addition, since the values of x_t are known, we can determine $g_1(x)$ as a linear or nonlinear function, not necessarily a polynomial, in the independent variable x. Such an analysis yields a determinate set of individual curves as polynomials h_r in the $g_1(x)$ so determined. That is, by the methods of sect. 3.2, and in addition the determination of $g_1(x)$ we obtain individual functions of the type

$$y_{it} = \sum_{r=1}^{s} b_{ir}h_r\{g_1(x)\} + e_{it} \tag{4.8.16}$$

in which the b_{ir} are estimated individual parameters, and the $h_r\{g_1(x)\}$ may be thought of as *reference curves.*

Method II: Procedure Using Classical Curve-Fitting and Linear Factor Analysis

There is, however, an alternative approach to the problem which is both simpler and, in some ways, more instructive. Following the classical procedures for curve fitting by orthogonal polynomials, we write

$$y_{it} = \sum_{j=0}^{\ell} k_{ij}h_j(x_t) + e_{it} \tag{4.8.17}$$

where the functions $h_j(x_t)$ are, as in (3.2.1), orthonormal polynomials in x of precise degree j, i.e.,

$$\sum_{t=1}^{m} h_j(x_t)h_k(x_t) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \tag{4.8.18}$$

and $\ell + 1 \leq \min(m, n)$.

We may rewrite (4.8.17) in matrix form,

$$Y = KH(x) + E, \tag{4.8.19}$$

* It may be noted that in contrast to Tucker's (1960) treatment, we are here developing the argument in terms of $\mathcal{E}(YY')$, i.e., interrelations between "persons" rather than "trials." This choice is virtually a matter of taste, since the preceding argument and the sequel could have been developed by introducing a transformation of A into an orthogonal matrix instead of the transformation T applied to $F(x)$. The argument here was originally developed on the basis of the account in Tucker (1958), in ignorance of the later work. Perhaps the only point to be made now about this choice is that the orthonormal functions in the independent variable x appear to have a more "substantive" basis than would orthonormal transformations of the parameters associated with the individuals in the set.

where $K \equiv [k_{ij}]$, an $n \times (\ell + 1)$ matrix, which is not necessarily of full rank, and

$$H(x) \equiv [h_i(x_t)], \quad \text{an } (\ell + 1) \times m \text{ matrix, with}$$

$$H(x)H'(x) = I. \tag{4.8.20}$$

By (4.8.4), (4.8.19) and (4.8.20),

$$\mathcal{E}(YY') = KK' + U^2, \tag{4.8.21}$$

but in this case K is not necessarily of full rank, in contrast to (4.8.14).

In case the x_t are equally spaced, the functions h_i can be obtained by normalizing the Tchebycheff polynomials listed by Pearson and Hartley (1956), and the elements of $H(x)$ are given by normalizing the corresponding tabulated values.

If $\mathcal{E}(e_{it}e_{iu}) = 0$ for all $t \neq u$ and all i, least-squares estimates $\hat{K}$, of K are given by

$$\hat{K} = YH'(x). \tag{4.8.22}$$

The usual tests of significance (Pearson and Hartley, 1956) can then be applied to successive elements of each row vector of $\hat{K}$. The use of these tests involves the further assumption that the joint density function of the errors is normal. Let us now suppose that as a result of such tests of significance, all column vectors of $\hat{K}$ for $j \leq p + 1$ $(p \leq \ell)$ require some nonzero entries, while all column vectors of $\hat{K}$ for $j > p + 1$ can reasonably be deleted. This procedure also yields estimates $\hat{U}^2$ of U^2.

Now if K is of rank $s \leq$ min. $(p + 1, n)$ we may write

$$K = MC^{1/2}N, \tag{4.8.23}$$

where M, N are orthogonal matrices, respectively, of orders $n \times s$ and $s \times (p + 1)$, and C is a diagonal matrix containing the s nonzero latent roots of KK'.

Then by (4.8.19) and (4.8.23)

$$Y = MC^{1/2}NH(x) + E \tag{4.8.24}$$

which we rewrite as

$$Y = MC^{1/2}T(x) + E, \tag{4.8.25}$$

where

$$T(x) = NH(x) \tag{4.8.26}$$

is an orthogonal matrix.

Hence if $\hat{K}$ is approximately of rank s, we may obtain estimates $\hat{M}$, $\hat{C}$, and $\hat{N}$ by the usual methods, and then write

$$Y = MC^{1/2}T(x) + E, \tag{4.8.27}$$

a set of individual curves expressed in terms of a minimum number of orthonormal polynomials, with the functions

$$T(x) = NH(x) \tag{4.8.28}$$

serving to describe the reference curves. (One simple procedure is to obtain the latent roots $\hat{C}$ and latent vectors $\hat{N}$ of $\hat{K}'\hat{K} = \hat{N}'\hat{C}\hat{N}$, which will in general be of much lower order than $\hat{K}\hat{K}' = \hat{M}\hat{C}\hat{M}'$, and then obtain M by

$$\hat{M} = \hat{K}\hat{N}'\hat{C}^{-\frac{1}{2}}.$$

Essentially, what this amounts to is that we first fit a set of individual curves by the classical methods, using the conventional tests of significance. We then use the ordinary methods of factor analysis to condense the matrix of individual parameters so obtained, according to the extent of linear dependence shown between these parameters, and apply the same condensation to the set of orthogonal polynomials. (As developed above, the argument does not allow for any alterations to the elements in the leading diagonal of KK', before obtaining its latent roots and vectors.)

Contrast, for example, the model

$$y_{it} = k_{i0} + k_{i1}x_t + k_{i2}h_2(x_t) + e_{it} \tag{4.8.29}$$

in which the parameters k_{i0}, k_{i1}, k_{i2} are linearly independent, with the model

$$y_{it} = k_{i0} + k_{i1}x_t + k_{i2}h_2(x_t) + k_{i3}h_3(x_t) + k_{i4}h_4(x_t) + e_{it} \tag{4.8.30}$$

in which the parameters are such that it may be rewritten in the form

$$y_{it} = c_{i0} + c_{i1}\{x_t + \alpha h_2(x_t)\} + c_{i2}\{h_3(x_t) + \beta h_4(x_t)\} + e_{it}, \tag{4.8.31}$$

where c_{i0}, c_{i1} and c_{i2} are linearly independent.

In the first of these models, the matrix of individual parameters is of full rank (rank three), the fitted individual curves are quadratic functions, and the application of the latent roots/latent vectors procedure will not yield any condensation of the expressions for the reference curves. In the second model, the matrix of individual parameters is also of rank three, but the fitted individual curves are quartics, and the application of the latent roots/latent vectors procedure yields some condensation in the expressions for the reference curves.

The application of Method I to data conforming to these models would yield similar expressions, but we might fail to realize the contrasting nature of the two cases.

CHAPTER 5

EMPIRICAL APPLICATIONS OF THE THEORY

5.0 *Introduction*

Each of the empirical studies reported in this chapter serves to illustrate one of the specializations of the theory that was given in Chap. 4. The sections are labeled in conformity with the sections of Chap. 4 which are here illustrated.

In sect. 5.1 two analyses of subtests of the Raven Progressive Matrices Test (1947) serve to illustrate the theory of difficulty factors given in sect. 4.1. In sect. 5.2 an analysis of items from two quasi-scales yields a two-factor latent polynomial for dichotomies. Section 5.3 contains applications of the treatment suggested in sect. 4.3 for latent class analysis. An attempt to apply the treatment of the latent distance model given in sect. 4.5 is reported in sect. 5.5. Finally, an application of the methods of curve fitting by factor analysis, given in sect. 4.8, to a set of learning curves is discussed in sect. 5.8.

5.1 *Two Analyses of Raven's Progressive Matrices Test* (1947)

Introduction

It has been shown by Gabriel (1954) that if the five subtests in the 1938 version of the Progressive Matrices Test are intercorrelated, the properties of the resulting correlation matrix are in reasonable accordance with Guttman's simplex theory. This test consists of five sets of twelve incomplete patterns, and the examinee is asked to complete each pattern by choosing one of six or eight alternatives. The items tend to increase in difficulty through each set, and the successive sets are claimed to increase in logical complexity, as well as difficulty.

In the theory presented above (sect. 2.3 and 4.1) it was pointed out that the factors obtained in the analysis of a simplex correlation matrix may be identical with the difficulty factors discussed by Ferguson and interpreted by Gibson as being due to curvilinear relations between tests and factors. The distinction with which we are concerned is between a model in which each test loads on one more factor than its predecessor in the sequence (i.e., the tests show a successive increase in factorial complexity) and a model in which the difficulty factors that emerge are components due to curvilinear regressions of the tests on a single factor. In the latter case

the differences in curvature correspond to the fact that easy tests will discriminate best at the lower end of the ability dimension, while the difficult tests discriminate best at the higher end.

From these considerations, it seems worthwhile to reanalyze the Progressive Matrices Test (PM) using the techniques of nonlinear factor analysis. For this purpose, however, there are two unsatisfactory features about the 1938 version of the test.

Firstly, subset *A* of the test appears to be rather too easy to be worth including in an analysis, and the remaining four subsets do not provide a large enough correlation matrix to be "fair" to the present theory. (It will be recalled that this theory rests in part on the operation of the central limit theorem as the number of observed variables becomes large.)

Secondly, the test as a whole is too easy to give a satisfactory spread in performance on a university population. In contrast, the 1947 "high-level" version of the test proves to give a satisfactory spread in performance both for university students and in a sample of 13-year-old children. Furthermore, it consists of forty-eight items, of which successive sets of four are claimed by Raven to constitute subtests of increasing complexity. (Yates (1961) found that two subsets are misplaced in difficulty level.) Hence it can be made to yield twelve subtests for the purpose of factor analysis.

Accordingly, it seemed desirable to carry out a nonlinear factor analysis on the twelve subtests of the 1947 PM, using both a sample from a university population and a sample of children.* One could hope to find in one or both analyses that the regressions of the tests on one factor would be nonlinear, and that the loadings on the quadratic component should correlate with test difficulty, as discussed in sect. 4.1. Further, one would hope to find that the regression curves for the sample of children would link up in a sensible way with the curves for the university sample. A priori, there were a number of possibilities. In particular, it seemed quite possible that the data might be in accord with Guttman's model, that is, a linear factor model with successive increases in factorial complexity.

Procedure

The 1947 PM was administered under virtually untimed conditions to groups of boys in Sydney high schools. Results were available from 390 subjects, mean age 13.96 years, SD 0.49 years. Of the total of 18,720 responses required, only twenty-six items were not attempted. The sample of children will be referred to as Group C. The test was also administered in a single session with a seventy-minute time limit to ninety-three beginning students in psychology, mean age 18.54 years, SD 2.32 years. Of the total of 4,464 responses required, seven items were not attempted. The sample of

* The data from the children were kindly supplied by Mr. R. A. Lockhart, Department of Psychology, University of Sydney.

TABLE 5.1.1c

Test	Mean	SD
1	3.677	0.663
2	3.303	1.075
3	2.210	1.135
4	2.659	1.157
5	2.310	1.304
6	1.154	1.135
7	1.897	1.062
8	1.415	1.070
9	1.687	1.267
10	0.562	0.744
11	0.544	0.692
12	0.331	0.546

ninety-three adults will be referred to as Group A. Results for the two groups will be presented in parallel, to facilitate comparisons. Parallel tables and graphs will be differentiated by a "c" or an "a" termination to identify the group from which they originate.

Results

Means and SDs of the twelve tests are given in Table 5.1.1. The sample correlation matrices are given in Table 5.1.2. In both cases, three latent roots were larger than unity in a preliminary principal component analysis with unities in the leading diagonals. Communality estimates were therefore obtained on the hypothesis of rank three for both matrices, and the principal component analysis was repeated using these estimates in the diagonals.

TABLE 5.1.1a

Test	Mean	SD
1	3.881	.353
2	3.892	.341
3	3.700	.598
4	3.690	.602
5	3.636	.581
6	3.009	1.207
7	3.030	.818
8	2.924	1.074
9	3.222	1.012
10	1.446	1.126
11	1.446	1.078
12	1.021	.911

TABLE 5.1.2c

Matrix R: Group C

(.276)	.362	.124	.333	.388	.253	.325	.307	.285	.048	.125	.007
.400	(.475)	.177	.442	.514	.338	.434	.384	.410	.085	.173	.017
.158	.114	(.287)	.232	.260	.223	.263	.276	.242	.383	.202	.126
.345	.444	.205	(.432)	.500	.345	.435	.398	.409	.187	.203	.052
.377	.577	.210	.476	(.578)	.396	.501	.457	.471	.202	.230	.055
.202	.298	.206	.345	.357	(.283)	.343	.332	.331	.214	.183	.064
.361	.375	.258	.429	.515	.347	(.442)	.412	.415	.240	.221	.071
.218	.383	.289	.433	.431	.325	.430	(.390)	.385	.280	.223	.086
.292	.343	.251	.410	.461	.395	.420	.398	(.389)	.216	.205	.063
.035	.169	.139	.190	.248	.174	.238	.277	.173	(.572)	.249	.193
.081	.158	.112	.190	.200	.272	.236	.246	.214	.249	(.150)	.081
.132	.004	.016	.089	.097	.030	.045	−.007	.072	.010	.074	(.065)

With these communality estimates, the two matrices of factor loadings are as shown in Table 5.1.3. The component variates $[v_1, v_2, v_3]$ were also estimated. Scatter diagrams of v_2 on v_1 are shown in Fig. 5.1.1.

In both cases the third component variate v_3 turned out to have a peculiar trimodal distribution, with clear separations between the three groups. It was discarded from the present analysis in both cases and is discussed below in sect. 5.4.

The factor loadings on the first two factors are graphed in Fig. 5.1.2. The upper triangles in Table 5.1.2 contain the reproduced correlations as

TABLE 5.1.2a

Matrix R: Group A

(.023)	.032	.073	.059	.074	.103	.084	.091	.084	.089	.073	.053
.515	(.046)	.100	.082	.075	.146	.114	.137	.119	.117	.126	.071
−.014	−.051	(.230)	.188	.233	.326	.266	.285	.266	.283	.224	.169
.130	.047	.158	(.154)	.192	.267	.217	.235	.218	.230	.189	.138
.054	.143	.203	.153	(.658)	.282	.298	.129	.242	.377	−.134	.216
.053	.132	.402	.327	.309	(.471)	.373	.430	.383	.388	.376	.233
−.006	−.086	.284	.307	.224	.333	(.307)	.321	.306	.330	.243	.197
.033	.153	.363	.228	.097	.403	.250	(.426)	.346	.316	.437	.193
.134	.193	.200	.108	.280	.333	.359	.354	(.312)	.319	.296	.192
.034	−.006	.274	.202	.319	.356	.377	.328	.294	(.362)	.203	.215
.013	.086	.147	.154	−.070	.396	.227	.354	.319	.186	(.562)	.131
−.025	.042	.129	.167	.255	.145	.169	.209	.139	.260	.208	(.128)

Lower triangle, observed correlations
Upper triangle, reproduced correlations (2 factors)
Diagonal, communalities (2 factors)

TABLE 5.1.3c

Matrix F: Group C

$$\begin{bmatrix} .476 & -.222 & .135 \\ .637 & -.264 & -.012 \\ .417 & .337 & -.124 \\ .648 & -.112 & .036 \\ .745 & -.150 & .039 \\ .532 & .002 & -.034 \\ .664 & -.040 & -.020 \\ .623 & .048 & -.103 \\ .622 & -.051 & .011 \\ .400 & .642 & -.195 \\ .344 & .174 & -.009 \\ .120 & .226 & .929 \end{bmatrix}$$

$C = \text{diag}\,[3.572 \quad .768 \quad .949]$

from the first two factors. The major discrepancies between observed and reproduced correlations are due to the omission of the third factor in each case. The joint moments of $[v_1, v_2]$ are given in Table 5.1.4, together with the moments of the "true" parts $[w_1, w_2]$.

From Fig. 5.1.1, it is reasonably evident in the case of Group C that the points $[v_1, v_2]$ cluster about a quadratic parabola whose axis is approximately vertical. The picture is not nearly so clear in the case of Group A, but with a little imagination one can project into it a quadratic parabola whose principal axis is rotated approximately 30 to 45 degrees anticlockwise from the vertical. In accordance with the theory given in sect. 3.1 and 3.2,

TABLE 5.1.3a

Group A

$$F \qquad\qquad B$$

$$\begin{bmatrix} .151 & -.006 & -.725 \\ .211 & .035 & -.803 \\ .479 & -.030 & .168 \\ .392 & -.014 & .000 \\ .467 & -.663 & -.055 \\ .684 & .057 & .038 \\ .551 & -.062 & .201 \\ .609 & .235 & .023 \\ .558 & .028 & -.090 \\ .580 & -.160 & .133 \\ .503 & .556 & .047 \\ .348 & -.080 & .066 \end{bmatrix} \qquad \begin{bmatrix} .134 & -.069 \\ .206 & -.058 \\ .421 & -.230 \\ .349 & -.178 \\ .143 & -.798 \\ .644 & -.238 \\ .473 & -.289 \\ .651 & -.045 \\ .517 & -.211 \\ .458 & -.390 \\ .691 & .291 \\ .281 & -.220 \end{bmatrix}$$

$C = \text{diag}\,[2.833 \quad .846 \quad 1.276]$

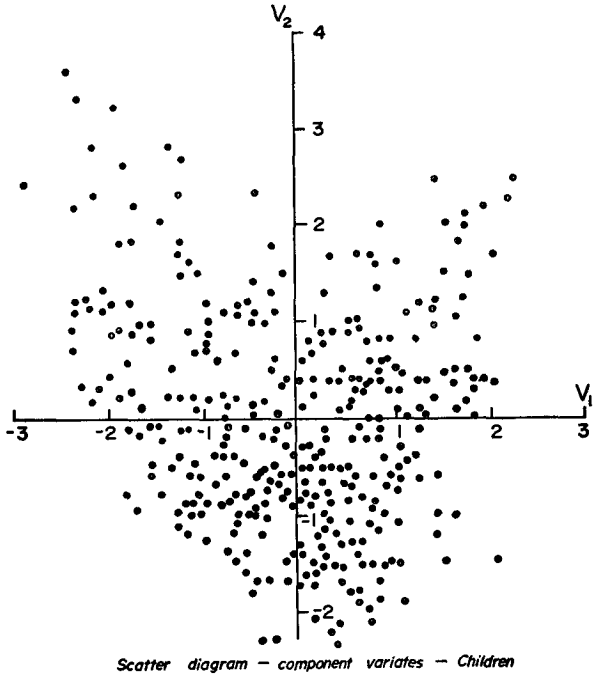


Figure 5.1.1c

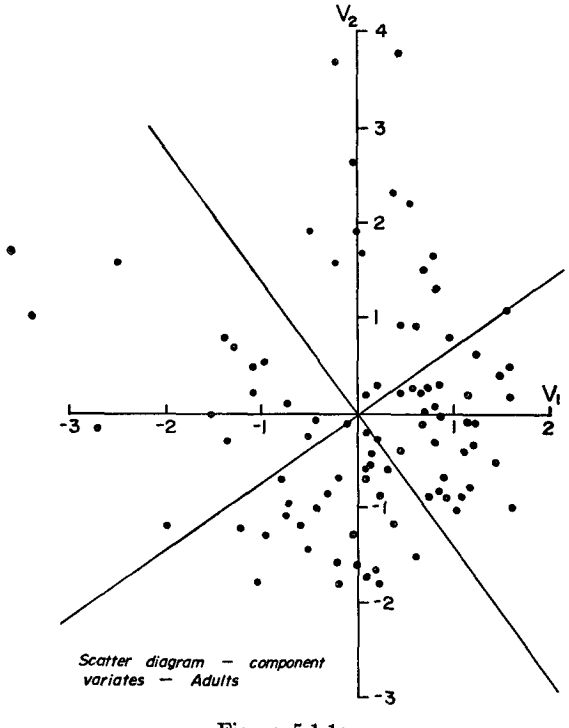


Figure 5.1.1a

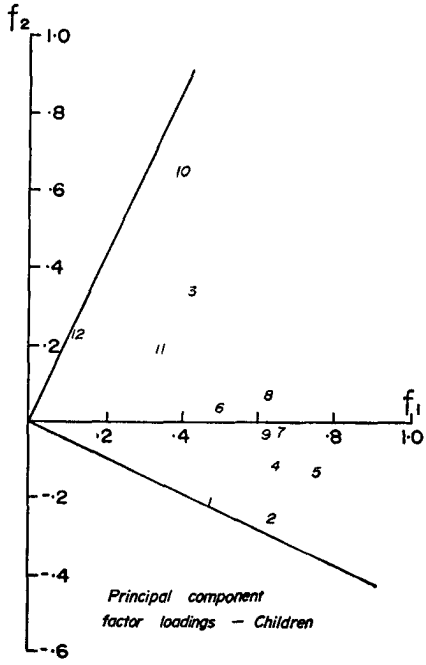


Figure 5.1.2c

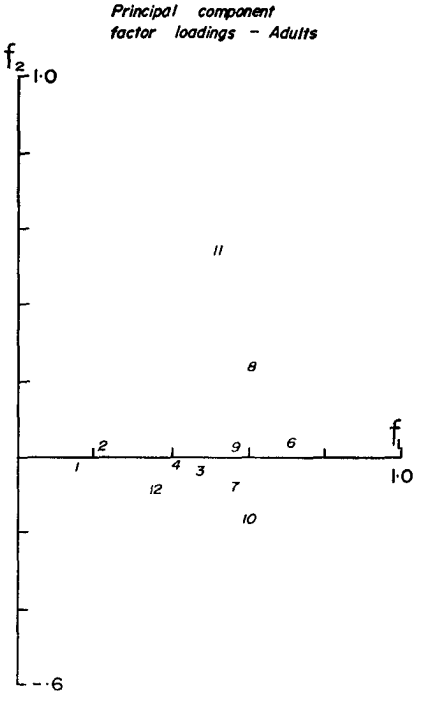


Figure 5 1.2a

TABLE 5.1.4c

Moments: Group C

μ	v	w
20	1.146	1.000
11	−0.125	0.000
02	1.352	1.000
30	−0.160	−0.160
21	0.774	0.774
12	−0.004	−0.004
03	0.601	0.601
40	3.232	2.292
31	−0.526	−0.096
22	2.314	1.765
13	−0.878	−0.371
04	5.500	3.016

we wish to find, for both cases, a rotation θ which will give a best parabola, in the sense described earlier, to fit these graphs.

In the case of Group C, Fig. 5.1.3 provides a graph of Φ_2 for θ in the range $-10°$ to $+10°$. The required rotation is approximately zero, hence we may be satisfied to take

$$k = 0.600, \qquad \mu_3 = -0.160, \qquad \theta = 0,$$

yielding

$$\Phi_2(\text{min}) = 0.536.$$

TABLE 5.1.4a

Moments: Group A

μ	v	w
20	1.221	1.000
11	.093	0.000
02	1.339	1.000
30	−.687	−.687
21	.640	.640
12	−.015	−.015
03	1.318	1.318
40	4.400	2.928
31	.048	−2.93
22	2.745	2.110
13	.225	−.149
04	6.075	3.696

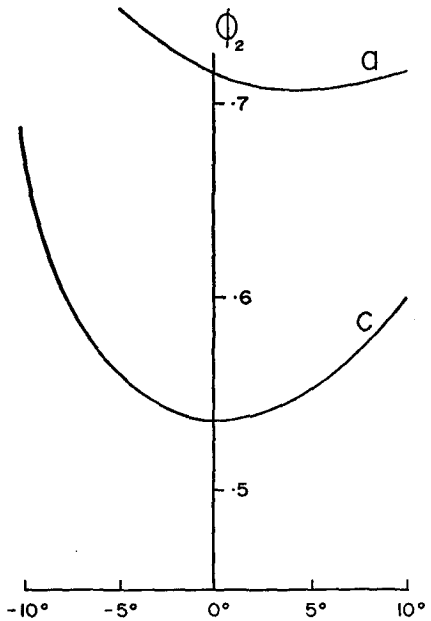


Figure 5.1.3

An analysis of the variance of v_2 into orthogonal components is given in Table 5.1.5c. The F of 112.832 on one and 387 df for the quadratic component seems adequate confirmation of the hypothesis that the "true" part w_2 is a quadratic function of w_1. Hence with zero rotation, we confirm the model $[x\ 0.6(x^2 + 0.16x - 1)]$ and reject the alternative hypothesis of two satistically independent factors with linear regressions. With the obtained expression for the quadratic term, the regression functions of the tests on a single factor x are as given in Table 5.1.6c. Table 5.1.7c gives the corresponding expressions when the test measurements are converted back into raw score form. A selection of these regression functions is graphed in Fig. 5.1.4c. There is an obvious similarity (with the notable exception of test 3) between these regressions and the curves in Fig. 4.1.1 discussed earlier. That is, the easy tests show maximum slope, corresponding to maximum discrimination, at the negative end of the x dimension, while the difficult tests show maximum slope at the positive end.

In the case of Group A, a preliminary graphical rotation of the points $[v_1, v_2]$ through an angle of $-30°$ was carried out on the $[v_1, v_2]$ graph as indicated. An analysis of the variance of v_2, after rotation, into orthogonal components was then performed to verify the rather imaginative hunch mentioned above. The results of this are given in Table 5.1.5a. The F of 15.03 on 1 and 90 df ($P < 0.001$) confirms the hypothesis that the "true"

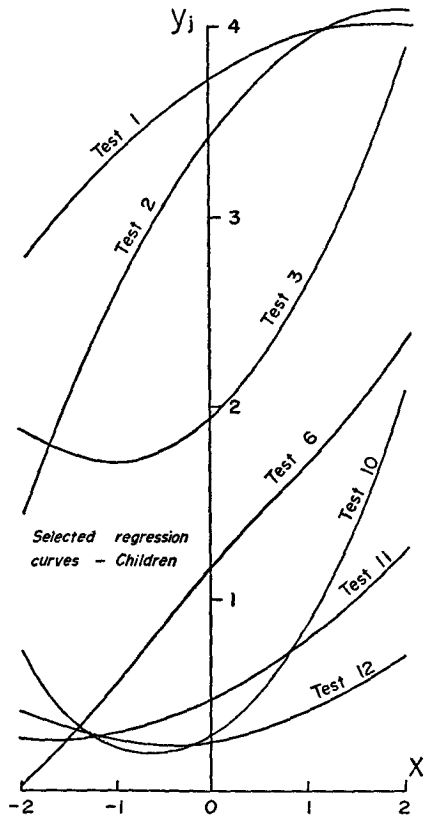


Figure 5.1.4c

TABLE 5.1.5c
Analysis of Variance: Group C

Source	SS	df	Ms	*F*
Total	527.280	389		
Linear	5.312	1	5.312	5.088*
Quadratic	117.797	1	117.797	112.832***
Residual	404.171	387	1.044	

TABLE 5.1.5a
Analysis of Variance: Group A

Source	SS	df	Ms	F
Total	124.20	92		
Linear	0.66	1	0.66	1
Quadratic	17.68	1	17.68	15.03***
Residual	105.86	90	1.176	

parts $[w_1, w_2]$ lie on a rotated quadratic parabola, as surmised on the basis of inspecting the graph. Fig. 5.1.3 provides a graph of Φ_2 for θ in the range $-10°$ to $+10°$ about the initial rotation of $-30°$. The minimum of Φ_2 with respect to θ is at $+5°$, hence the required total rotation is through $-25°$, and the required parameters are

$$k = 0.425, \qquad \mu_3 = -0.511, \qquad \theta = -25°,$$

yielding

$$\Phi_2(\text{min}) = 0.707.$$

TABLE 5.1.6
Estimated Parameters: Standard Score Form
$z = c_{1j}x - c_{2j}(x - 1)$

Test	Group C		Group A	
	c_{j1}	c_{j2}	c_{j1}	c_{j2}
1	.455	−.133	.099	−.029
2	.612	−.158	.176	−.025
3	.449	.202	.303	−.098
4	.637	−.067	.258	−.076
5	.731	−.090	−.265	−.339
6	.532	.002	.522	−.101
7	.660	−.024	.325	−.123
8	.628	.029	.628	−.019
9	.617	−.031	.409	−.090
10	.462	.385	.259	−.166
11	.361	.104	.840	.124
12	.142	.136	.169	−.093

TABLE 5.1.7
Estimated Parameters: Raw Score Form
$y = a_i + b_i x + c_i x$

Test	Group C			Group A		
	a_i	b_i	c_i	a_i	b_i	c_i
1	3.755	.302	−.088	3.891	.035	−.010
2	3.473	.658	−.170	3.901	.060	−.009
3	1.981	.510	.229	3.759	.181	−.059
4	2.737	.737	−.078	3.736	.155	−.046
5	2.427	.953	−.117	3.833	−.154	−.197
6	1.152	.604	.002	3.131	.630	−.122
7	1.922	.701	−.025	3.131	.266	−.101
8	1.384	.672	.031	2.944	.674	−.020
9	1.726	.782	−.039	3.313	.414	−.091
10	0.276	.344	.286	1.633	.292	−.187
11	0.472	.250	.072	1.312	.906	.134
12	0.257	.078	.074	1.106	.154	−.085

Correspondingly, a rotation matrix

$$L = \begin{bmatrix} .906 & -.423 \\ .423 & .906 \end{bmatrix}$$

is applied to the factor matrix of Table 5.1.3a (first two column vectors only). With the obtained expression $0.425(x^2 + 0.511x - 1)$ for the quadratic term, and the obtained matrix B of rotated factor loadings, we have the regression functions of the tests on the single factor x as given in Table 5.1.6. Table 5.1.7 gives the corresponding expressions when the test measurements are converted back into raw score form. A selection of these regression functions is graphed in Fig. 5.1.4a.

In Group C, the correlation between the test means and their loadings on the second factor is -0.715 ($t = 3.235$ on 10 df, $P < 0.01$). In contrast, in Group A, the corresponding correlations both before rotation ($r = -0.289$) and after rotation ($r = -0.206$) are nonsignificant.

Test 3 in Group C and test 5 in Group A are both represented as having markedly nonmonotonic regression functions in the region where there is appreciable population density. These were examined item by item to see if they might be genuinely nonmonotonic. A fairly independent check is given by plotting the proportion of correct responses as a function of total test score. None of the items appeared nonmonotonic on this basis.

Finally, we consider the relationship between the sets of regression curves obtained from the two samples. A general theory of factorial invariance

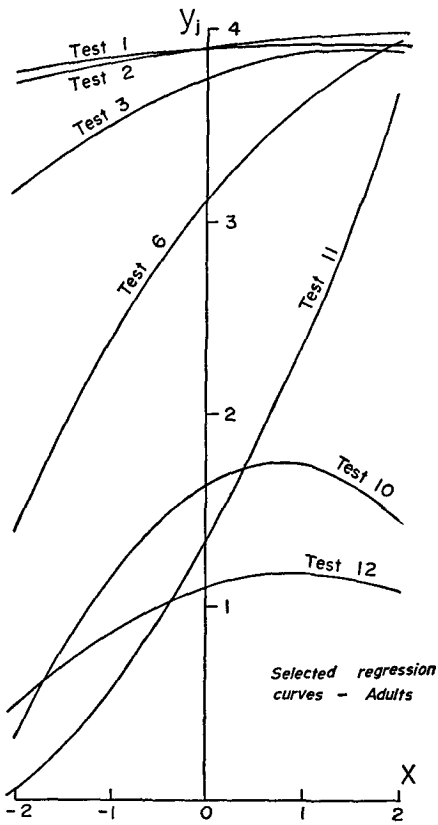


Figure 5.1.4a

that will be applicable to nonlinear factor analysis has not yet been worked out. For the purposes of the present example, it seemed sufficient to determine the points on the x-continuum for each group where the regression curves showed maximum agreement in the sense of least squares, and to examine the extent of the agreement at this point.

Writing

$$y_{jc} = a_{jc} + b_{jc}x_c + c_{jc}x_c^2$$

for the regression curves for Group C and

$$y_{ja} = a_{ja} + b_{ja}x_a + c_{ja}x_a^2$$

for the curves for Group A, we minimize the quantity

$$(5.1.1) \qquad \sum_j (y_{jc} - y_{ja})^2 = x_c^4 \sum_j a_{jc}^2 + x_a^4 \sum_j a_{ja}^2 + x_c^2 \sum_j b_{jc}^2 + x_a^2 \sum_j b_{ja}^2$$

$$+ \sum_j c_{jc}^2 + \sum_j c_{ja}^2 - 2x_c^2 x_a^2 \sum_j a_{jc}a_{ja} + 2x_c^3 \sum_j a_{jc}b_{jc}$$

$$+ 2x_a^3 \sum_j a_{ja}b_{ja} - 2x_c^2 x_a \sum_j a_{jc}b_{ja} - 2x_c x_a^2 \sum_j a_{ja}b_{jc}$$

$$+ 2x_c^2\left(\sum_j a_{jc}c_{jc} - \sum_j a_{jc}c_{ja}\right) + 2x_a^2\left(\sum_j a_{ja}c_{jc} - \sum_j a_{ja}c_{ja}\right)$$

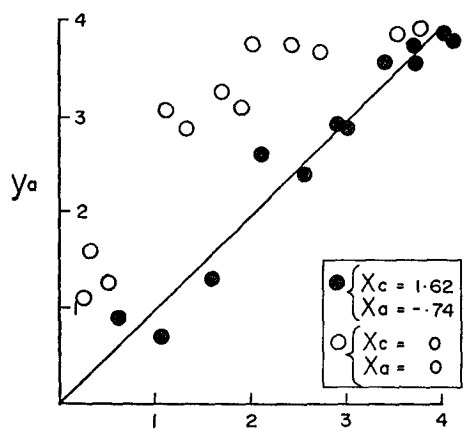
$$- 2x_c x_a \sum_j b_{jc}b_{ja} + 2x_c\left(\sum_j b_{jc}c_{jc} - \sum_j b_{jc}c_{ja}\right)$$

$$+ 2x_a\left(\sum_j b_{ja}c_{ja} - \sum_j b_{ja}c_{jc}\right) - 2\sum_j c_{jc}c_{ja},$$

with respect to x_c, x_a, using the values of the parameters given in Table 5.1.7. This yields a well-defined minimum value of 0.741 at $x_c = 1.62$, $x_a = -0.74$. The extent of misfitting of the individual regression curves can be seen in Fig. 5.1.5, which gives the values of the Group A regression functions at $x_a = -0.74$ plotted against the values of the Group C regression functions at $x_c = 1.62$. Unfortunately, there does not seem to be any test which could be applied in order to determine whether the agreement is "satisfactory" in some sense.

Discussion

In general, the regression functions relating the tests to the factor are plausible in both cases, thus serving to show that the methods of nonlinear



Relations between regression curves for Adults and Children at the point of best fit. Relation at $X_c = 0$, $X_a = 0$ *shown for contrast.*

Figure 5.1.5

factor analysis presented in Chap. 3 can be put to work and that empirical data can be found to which these methods are applicable.

The regression functions are "plausible" in that the following remarks apply. With the two exceptions noted above they are monotonic increasing functions of x in the region where there is appreciable population density (remembering that x is in standard form, hence most of the population will lie in the range $-2 < x < 2$, say). Further, in Group C, as noted above, the easy tests have their greater slope at the low end of the range of ability, while the difficult tests have their greater slope at the high end of the ability range, much as we would expect. In this respect, for the case of Group A, nearly all the tests behave like "easy" tests, with maximum discrimination at the lower end of the ability range. This could be expected in a university population as contrasted with school children.

The curves for the two groups show maximum agreement at $x_c = 1.62$, $x_a = -0.74$, and the amount of agreement appears to be satisfactory. It seems plausible that a child performing at 1.62 standard deviations above the mean of his group should have the same ability as a university student performing at 0.74 standard deviations below the mean of his group.

An interesting feature of this particular example is that it provides us, for the first time, with direct evidence that a difficulty factor can be due to differential curvature of the regressions on the "content" factor, together with evidence that it is possible to have differential curvature without a difficulty factor.

In the case of Group C, investigators employing linear factor analysis could be expected to notice that the loadings on the second factor correlate highly with test difficulty, and hence to set it aside as a difficulty factor, requiring no interpretation. The present methods serve to show further that in this case the "difficulty factor" corresponds quite directly to the shape of the regression curves. Also, the treatment by nonlinear methods gives us curves from which the discriminating power of each test, measured by the slope of the regression function, can be determined at any part of the ability range. In this way, the "difficulty factor," instead of being a nuisance, acquires some usefulness.

In the case of Group A, linear factor analysis might be expected to go quite astray in its treatment of the data. Here there is nothing to lead one to suspect a difficulty factor. In Fig. 5.1.2a one might follow the usual conventions for analyzing cognitive tests, and rotate the axes to yield a positive manifold. This is approximately attainable. (Simple structure cannot be remotely approached in this case.) Test 5 and test 11 then provide the only "pure" measures of the two alleged factors, other tests involving mixtures of these. One might even achieve an interpretation of these. But, from what we now know, the whole of this rotation and interpretation procedure would clearly be quite wrong, from start to finish. In fact, the rotation required by

the nonlinear procedures has nothing to do with positive manifold—in this case it produced a majority of negative loadings on the second factor—and certainly nothing to do with the principles of simple structure.

The Group C data required no rotation in that there was a close balance between positive and negative curvatures in the set of regression curves, hence the principal component solution was satisfactory as it stood. In the case of Group A, rotation from the principal components solution was necessary in that most of the regression curves showed curvature in the one direction.

5.2 *An Empirical Example of a Polynomial in Two Factors from Dichotomous Items*

Introduction

As pointed out in sect. 4.2 the latent polynomial model as described by Lazarsfeld is a model for observed dichotomies in which the regression functions are polynomials in a single latent variate or factor. If the model is appropriate, it can be fitted to empirical data by the methods of sect. 3.2.

Somewhat more generally, analysis of empirical data may demand a model in which the regression functions are polynomials in two or more latent variates. The methods of sect. 3.3 are then appropriate.

When looking for a set of dichotomous data that should serve to illustrate this type of case, it seemed advisable to obtain a reasonably large number of items which would have a fair a priori probability of requiring a nonlinear model. An obvious way to get some assurance of this is to use items which are known to form a quasi-scale. It is, however, unusual to find quasi-scales based on a reasonably large number of items.

Schutz (1958) has developed an Interpersonal Relations Inventory, consisting of six subtests containing nine items each. Each of the six subtests forms a fairly good quasi-scale. Data were available on this test for 158 subjects.* The six subscales are not mutually orthogonal. The six scales are made up of three content areas, viz. "inclusion" (in groups), "control" (over others), and "affection," by two aspects of each, "wanted behavior" (with items beginning "I like. . ." or equivalently) versus "expressed behavior" (with items beginning "I try to. . ." or equivalently). The highest intercorrelations between total scores on different scales are +0.62 between wanted and expected inclusion, and +0.57 between wanted and expected affection. The scatter diagram for the second of these relations revealed that peculiar but not uncommon form of distribution in which all the sample density was concentrated in one triangle, corresponding to a tendency for a score of y on one scale to be conditional on a score of greater than y on the other, but

* These data were kindly supplied by Mr. G. Singer of the Department of Psychology, University of Sydney.

TABLE 5.2.0

Schutz Scale

Item No.	
1	I try to be friendly to people.
2	My personal relations with people are cool and distant.
3	I act cool and distant with people.
4	I try to have close relationships with people.
5	I try to have close, personal relationships with people.
6	I try to have close relationships with people.
7	I try to get close and personal with people.
8	I try to have close, personal relationships with people.
9	I try to get close and personal with people.
10	I like people to act friendly toward me.
11	I like people to act cool and distant toward me.
12	I like people to act distant toward me.
13	I like people to act cool and distant toward me.
14	I like people to act distant toward me.
15	I like people to act close toward me.
16	I like people to act close and personal toward me.
17	I like people to act close and personal with me.
18	I like people to act close toward me.

Note.—(1) Items are scored positively or negatively where appropriate. (2) Items have one of two sets of response alternatives: (a) most people—many—some—a few—nobody; (b) usually—often—sometimes—occasionally—never. Where items are repeated, the "time" and the "people" alternatives are used. Cutting-points also vary.

not vice versa. These preliminary observations suggested that it might be interesting to factor the eighteen items for "wanted" and "expressed affection" put together in a single score matrix. These items are listed in Table 5.2.0. It may be noted that the items within each subscale reveal a kind of "semantic triviality," in that one finds virtually the same content from one item to the next, with a variant on the wording designed primarily to alter the probability of endorsing it. This seems rather typical of Guttman scales.

In the other examples in this chapter, there is in each case some kind of "validation" for the method, e.g., the relation of the second factor to test difficulty in the Group C data in sect. 5.1. In this case the analysis is best regarded as purely exploratory, and the results can only be judged in terms of their intrinsic plausibility.

Analysis and Results

The coefficients of reproducibility for the two scales employed were, respectively, 0.942 for "expressed affection" and 0.938 for "wanted affection," for 158 subjects.

In this, and most succeeding analyses, we are concerned with dichotomous

observed variables. Hence it is appropriate to factor the covariance matrix, in accordance with the argument of sect. 2.3, in order that the obtained regression functions will be directly interpretable as "trace functions," i.e., as representing the conditional probability, given the latent characterization, of giving the positive response to each item. Corresponding to the Guttman lower bound for the rank of the reduced correlation matrix, we require a criterion for the rank of the reduced covariance matrix. An approximate equivalent to the Guttman lower bound is obtained by taking the mean of the variances of the observed variables (the trace of the covariance matrix divided by its order). Then the rank of the reduced covariance matrix can be taken as equal to the number of latent roots of the original covariance matrix which are greater than the mean of the variances.

The observed covariances are given in Appendix 5.2A. The item means (p_i, the proportion of positive responses) and variances ($p_i(1 - p_i)$) are given in Table 5.2.1. In the initial principal components analysis, three latent roots (0.763, 0.501, and 0.196) were greater than the mean of the variances (0.151). Accordingly, the principal components analysis was repeated with the communalities obtained from these three factors, yielding the factor matrix and latent roots given in Table 5.2.2. This F matrix yields the reproduced covariances and communalities given in the upper triangle and leading diagonal of Appendix 5.2A.

TABLE 5.2.1

Item	F			p_i	p_iq_i	Communalities	Comm./Variance
1	.151	−.076	.018	.874	.1100	.0289	.263
2	.209	−.238	.076	.723	.2002	.1061	.530
3	.230	−.191	.023	.679	.2179	.0899	.413
4	.218	.207	−.179	.245	.1851	.1224	.661
5	.092	.135	−.119	.119	.1100	.0409	.372
6	.180	.139	−.018	.138	.1192	.0520	.436
7	.125	.138	.000	.094	.0905	.0347	.383
8	.121	.113	.047	.132	.1146	.0296	.258
9	.078	.091	−.059	.057	.0478	.0178	.372
10	.105	−.040	−.006	.931	.0644	.0127	.197
11	.222	−.113	−.098	.818	.1491	.0717	.481
12	.170	−.080	−.067	.780	.1717	.0398	.232
13	.297	−.168	−.040	.748	.1883	.1180	.627
14	.333	−.193	.022	.686	.2132	.1486	.697
15	.252	.199	.168	.314	.2156	.1313	.609
16	.197	.206	.016	.208	.1645	.0815	.495
17	.186	.185	.159	.195	.1570	.0941	.599
18	.158	.106	−.052	.195	.1570	.0389	.248

C = diag {.6949 .4338 .1322}

TABLE 5.2.2

Analysis of Variance of v_2

Source	SS	df	Ms	F
Total	183.870	157		
Linear	0.009	1		<1
Quad.	85.503	1	85.503	135.504***
Resid.	98.358	155	0.631	

Analysis of Variance of v_3

Total	157	228.867	F
Between Groups	7	9.215	1.316 < 1
With	150	219.652	1.464

(v_3 observations grouped into eight classes by partitioning the v_1 axis)

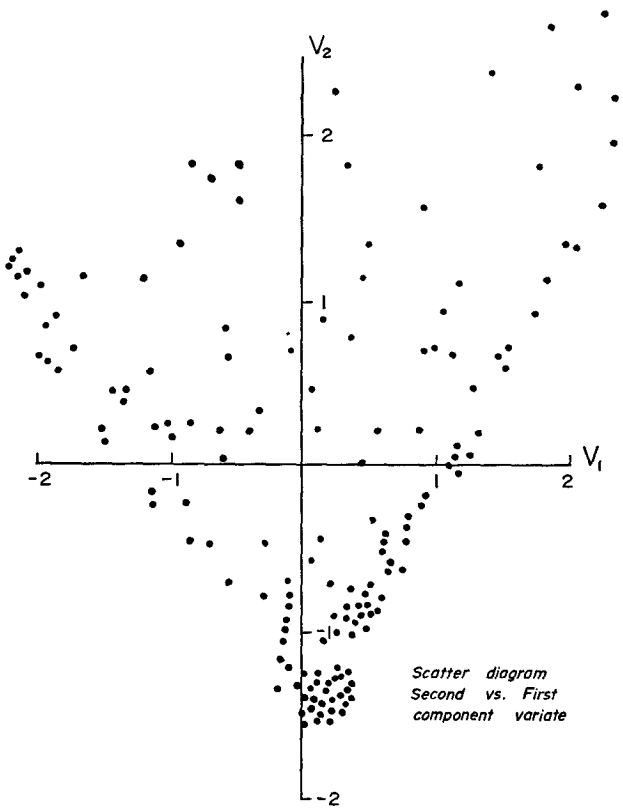


Figure 5.2.1a

Scatter diagrams of the distributions $[v_1, v_2]$ and $[v_1, v_3]$ are given in Fig. 5.2.1. Inspection of these suggests that the true part (w_2) of the second component variate can be expressed as a quadratic in w_1, with practically zero rotation, while v_3 appears to be statistically independent of v_1. Analyses of variance for v_2 and v_3 given in Table 5.2.2 confirm these impressions. Hence in this case we have the factor model

$$y_i = b_{0i} + b_{1i}x_1 + b_{2i}h_2(x_1) + b_{3i}x_2 + u_ie_i$$

where, as usual, $h_2(x_1)$ represents a quadratic function in x_1.

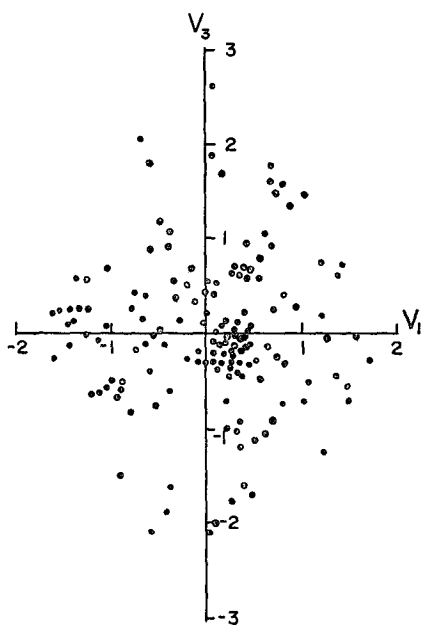
The required moments of the distribution of $[v_1, v_2]$ and the corresponding moments of the true parts $[w_1, w_2]$ are given in Table 5.2.3. Substituting the latter in the expression for Φ_2 as previously yields a required rotation of less than one degree, which may be safely neglected, and parameters k and μ_3 such that

$$h_2(x_1) = 0.155x + 0.629(x^2 - 1),$$

with

$$\Phi_2(\min) = 0.317.$$

Substituting the expression for $h_2(x_1)$ in the specification equations for the eighteen variables gives the regression functions (trace surfaces) of Table



Scatter diagram – Third vs. first component variate

Figure 5.2.1b

TABLE 5.2.3

Moments of (v_1, v_2) and (w_1, w_2)

μ	v	w
20	1.097	1.000
11	−.008	.000
02	1.154	1.000
30	−.247	−.247
21	1.085	1.085
12	.409	.409
03	.631	.631
40	3.396	2.785
31	.268	.293
22	.540	.274
13	.886	.913
04	3.040	2.045

5.2.4. Fig. 5.2.2 gives the regression curves for selected items on the first latent variate x_1.

Since in this example we require two statistically independent factors, x_1 and x_2, to account for the data, it is necessary to consider the alternative solutions that might be obtained by a transformation of $[x_1, x_2]$. A plot of the factor loadings on the third factor against those on the first is given in Fig. 5.2.3. From this it can be seen that it would not be possible, in the terms of conventional factor analysis, to rotate the solution to orthogonal simple structure. It is possible, however, by applying a rotation of $\sin^{-1} 0.8$ to achieve a positive manifold. Applying the transformation

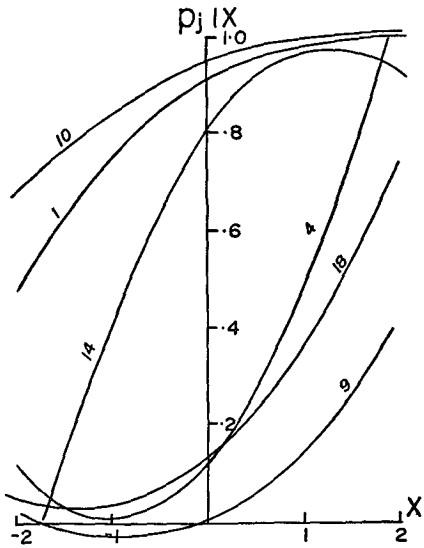
$$[\xi_1, \xi_2] = [x_1, x_2]\begin{bmatrix} .6 & .8 \\ -.8 & .6 \end{bmatrix}$$

to the expressions given in Table 5.2.4 gives the transformed regression functions of Table 5.2.5 (only a selection from the eighteen items is given).

Discussion

The regression functions obtained are fairly plausible, in that, over the range of x_1 for which the population density is not negligible, they give values for the probability of a positive response that lie for the most part within the required range of zero to one. The worst case is item 6, which is given a probability of endorsement of -0.064 at $x_1 = -1.0$. As in the examples of sect. 5.1, there are curves which are nonmonotonic within the range of the latent distribution, but none have a stationary value within the range $-1.0 < x_1 < 1.0$ which contains the major part of the population density.

Since in this case we require two factors to account for the data, ques-



Trace - functions of selected items
$p_j|X$ = *Probability of endorsement of item* j *for given* X
X = *factor score*

Figure 5.2.2

TABLE 5.2.4
Regression Functions
$\hat{y} = a_0 + a_{10}x_1 + a_{20}x_1^2 + a_{01}x_2$

Item	a_0	a_{10}	a_{20}	a_{01}
1	.922	.139	−.048	.018
2	.873	.172	−.150	.076
3	.799	.200	−.120	.023
4	.115	.250	.130	−.179
5	.034	.113	.085	−.119
6	.051	.202	.087	−.018
7	.007	.146	.087	.000
8	.061	.139	.071	.047
9	.000	.092	.057	−.059
10	.956	.099	−.025	−.006
11	.889	.204	−.071	−.098
12	.830	.158	−.050	−.067
13	.854	.271	−.106	−.040
14	.807	.303	−.121	.022
15	.189	.283	.125	.168
16	.078	.229	.130	.016
17	.079	.215	.116	.159
18	.128	.174	.067	−.052

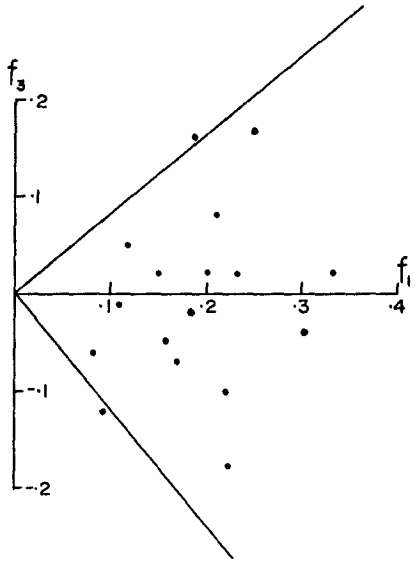


Figure 5.2.3

tions arise concerning rotation to a preferred solution and interpretation of the factors.

A comparison of the functions in Tables 5.2.4 and 5.2.5 serves to show that rotation of axes to achieve positive manifold, i.e., to a position such that all the first-order coefficients of ξ_1 and ξ_2 are positive, can only be achieved at the price of considerable complications to the description of the regression surface. This would still have been the case, though not quite to the same extent, if orthogonal simple structure were possible for the first-order coeffi-

TABLE 5.2.5

Regression Functions after Rotation

$$y = a_0 + a_{10}\xi_1 + a_{01}\xi_2 + a_{20}\xi_1^2 + a_{02}\xi_2^2 + a_{11}\xi_1\xi_2$$

Item	a_0	a_{10}	a_{01}	a_{20}	a_{02}	a_{11}
1	.922	.069	.122	−.017	−.031	.046
5	.034	.163	.019	.031	.054	−.082
9	.000	.102	.038	.021	.036	−.055
10	.956	.064	.076	−.009	−.016	.024
14	.807	.164	.256	−.044	−.077	.116
18	.128	.146	.108	.024	.043	−.064

cients. Further research experience with nonlinear factor analysis would seem necessary before any positive recommendations can be made with regard to this type of rotation problem.

There does not seem to be a clear-cut interpretation of the two factors that emerge from the analysis. Examining the unrotated form of the factor matrix in Table 5.2.1, one can regard the first, general factor, as concerned with a general orientation towards having close (affectionate?) relations with people. The third set of factor loadings should provide the necessary contrasts for the interpretation of the second latent variate. Here it is intriguing to note that the contrast, broadly, is between the items of one subscale with high probability of endorsement, together with the items of the other with low probability, as against the reverse combination. More specifically, it seems to reduce to a rather unlikely contrast of "trying to be close to people and wanting them not to be distant," versus "trying not to be distant from people and wanting them to be close." Beyond this semantic curiosity it seems impossible to go.

5.3 *Empirical Work Related to the Latent Class Model*

Introduction

In sect. 4.3 we saw that the methods employed as a matter of routine in nonlinear factor analysis can serve to reveal a genuine point distribution of the latent variate or factor, if such exists, and that in optimal cases, at least, it may be possible to estimate the parameters of the appropriate latent class model from the results obtained in the initial analysis.

It seems reasonable to suggest that latent class cases will prove to be comparatively rare in behavioral nature, while latent continuity would be the general rule. There is no evidence in the literature on this point, since in contrast to the procedures suggested in sect. 4.3, the methods of analysis given by Green (1951), Anderson (1954) and Gibson (1955) do not provide a test of the assumption that the latent distribution is discrete.

Thus, in an interesting application of latent class analysis to listeners' choices of radio programs, Gibson (cf. Lazarsfeld, 1959) was able to account for the data in terms of six latent classes. Four of these could be interpreted readily as: A, "non-listeners"; D, "high-brows"; E, "low-brows"; and F, "older and/or small-town people." Each of these verbal labels refers to an attribute which admits of degrees of variation. One can therefore suspect that an alternative model allowing for continuous variation in a latent space of one or more dimensions might be preferable, in comparison with a model which imposes a discrete class interpretation from the start.

Similarly, Miller, Sabagh and Dingman (1962) have applied latent class analysis, using Green's method, to five diagnostic characteristics of patients admitted to a hospital for the mentally retarded. The five variables were

adequately accounted for by a latent dichotomy; the two classes were interpreted respectively as "mildly retarded" and "severely retarded." In this case the interpretation suggests strongly that we should really have latent continuity, rather than a two-valued characteristic of the patient, "mild" versus "severe" retardation.* The complete set of data in this study was kindly made available by Dr. Miller for reanalysis.

Since the present program of research was begun, no really worthwhile example of a latent point distribution has been discovered. For the purposes of this section, then, two empirical cases will be presented. The first is a reanalysis of the data of Miller, Sabagh and Dingman, while the second consists of a further examination of the third factor in each of the two analyses given in sect. 5.1.

A Reanalysis of the Data of Miller et al.

The data as supplied by Dr. Miller are given in Table 5.3.1. This yields in turn the covariance matrix in Table 5.3.2. The estimates of the parameters of the latent dichotomy obtained by Miller et al. using Green's method are also given in Table 5.3.2. Since the reduced covariance matrix is approximately of rank one, it can equally be accounted for by a latent dichotomy or by the latent linear model. The estimates of the parameters a_j, b_j in a latent linear interpretation of the data are also given in Table 5.3.2. In the columns $a_j(s)$, $b_j(s)$, these estimates are obtained directly from the data, with the proportion of positive responses to each item as estimates of a_j, and with the slope parameters b_j obtained as factor loadings in a Spearman factor analysis of the covariance matrix, as recommended by Torgerson (1958). In the columns $a_j(M)$, $b_j(M)$, these parameters have been calculated from the latent class parameters, using the formulae (4.3.5) and (4.3.6) of sect. 4.3. Such small discrepancies as occur between the results of the two methods would be due, basically, to differences in the communalities obtained in the Spearman analysis and the equivalents to communality obtained in a latent class analysis by Green's method.

The value of the component variate, or, in this case, the *general factor score*, was calculated for each response pattern, using the factor loadings from the Spearman analysis of the covariance matrix. These values are given in Table 5.3.1. The distribution of v is represented as a histogram in Fig. 5.3.1. Considering the large size of the sample (3709 cases), this strongly suggests the existence of four believable modes, separated by three intervals of low frequency. One is tempted to say that the frequencies of the thirty-two response patterns might be accounted for by four latent classes on a unidimensional continuum. In terms of the theory given by Lazarsfeld (1950),

* The objects of Miller et al. were to study differential mortality rates as a function of the diagnostic characteristics employed. A decision as to the "genuineness" of the discontinuity imposed in the analysis is almost certainly unnecessary from their point of view.

TABLE 5.3.1

Pattern	Frequency	v
11111	314	1.766
11110	85	.748
11101	61	.660
11100	27	−.358
11011	49	1.296
11010	21	.278
11001	12	.190
11000	4	−.828
10111	200	.819
10110	112	−.199
10101	135	−.287
10100	274	−1.305
10011	38	.349
10010	37	−.669
10001	27	−.757
10000	158	−1.775
01111	416	1.688
01110	124	.670
01101	116	.582
01100	69	−.436
01011	68	1.218
01010	24	.200
01001	5	.112
01000	12	−.906
00111	218	.741
00110	125	−.277
00101	177	−.365
00100	374	−1.383
00011	50	.271
00010	44	−.747
00001	44	−.835
00000	289	−1.853

Variables

1		2		3		4		5	
	Sex		Age		Race		IQ		Dx
1	female	1	0–9 yrs.	1	Caucasian	1	0–39	1	Somatic
0	male	0	10 yrs.	0	Nonwhite	0	40	0	Nonsomatic

this would be a "degenerate" four-class case, degenerate in the sense that four classes are required, yet the covariance matrix is of rank one. However, before we take this notion too seriously, it is advisable to examine the response patterns more closely.

In Table 5.3.3a, the thirty-two response patterns have been put into

TABLE 5.3.2

	Covariance Matrix				Latent Class Parameters φ_{j1}	φ_{j2}	$a_j(s)$	$a_j(M)$	$b_j(s)$	$b_j(M)$
1					.400	.447	.419	.418	.016	.023
2	−.005				.153	.723	.379	.379	.265	.280
3	.007	.038			.650	.933	.762	.763	.136	.139
4	.014	.100	.035		.255	.924	.519	.520	.333	.327
5	.007	.084	.045	.095	.277	.892	.520	.521	.305	.301

g_1.604 g_2.396

TABLE 5.3.3a

Response Pattern Frequencies Grouped by v-value

(1)	(2)	(3)	(4)
00000(289)	00101(177)	00111(218)	01111(416)
10000(158)	01100(69)	01101(116)	11111(314)
00100(374)	10101(135)	01110(124)	
10100(274)	10110(112)	10111(200)	
00110(125)	11100(27)	11101(61)	
		11110(85)	

(1–2)	(2–3)	(3–4)
00001(44)	00011(50)	01011(68)
00010(44)	01001(5)	11011(49)
01000(12)	01010(24)	
10001(27)	10011(38)	
10010(37)	11001(12)	
11000(4)	11010(21)	

TABLE 5.3.3b

Response Pattern Frequencies
(Omitting Variables 1 and 3)

(1)	(2)	(3)	(4)
000(1095)	001(312)	011(418)	111(730)
010(125)	100(96)	101(177)	
	010(135)	110(209)	

(1–2)	(2–3)	(3–4)
001(71)	011(88)	111(117)
010(81)	101(17)	
100(16)	110(45)	

four groups corresponding to the intervals of v containing a high probability density, separated by the three groups corresponding to intervals of low density. Inspection of this table suggests that we consider what remains if we omit the first and third items (sex and ethnic status) which have proved, not surprisingly, to be poor discriminators between the two alleged latent classes. Omitting these items we obtain the corresponding set of response patterns given in Table 5.3.3b. From this table it appears that the intervals of high density in the histogram of Fig. 5.3.1 correspond fairly well to the total scores 0, 1, 2, and 3 of positive responses, in all possible patterns, for the three items which show good discrimination. (The behavior of the fourth item, IQ—second in the reduced response patterns—is exceptional. The 010 pattern turns up in three of the intervals as in Table 5.3.3b. Nevertheless, the overall picture seems clear enough.) Thus, in a real yet quite trivial sense, we have in this case four latent classes on a unidimensional continuum. On the whole, then, it would seem preferable to describe the data in terms of the latent linear model. A continuous gradation from mild to severe retardation makes psychological sense. At the same time it must be recognized that one cannot positively assert latent continuity on the basis of a small number of observed variables.

Empirical Example 5.1 *Revisited*

In the two analyses of the Raven PM data of sect. 5.1, a curious thing was noted. Three factors were retained in the initial analysis in each case and the distributions of component variate v_3 were multimodal, as shown in Table 5.3.4. It looked as though we would require two latent variates to account for the data in both cases. The first would be the variate appearing in the analyses given above, with quadratic regression lines. The second would be a variate taking only two or three discrete values as suggested by Table 5.3.4.

However, in the case of Set C, inspection of the factor loadings on the third factor revealed that test 12 had a loading of 0.929, while no other test had a loading of greater than ± 0.2 (cf. Table 5.1.3c). Since the scores on test 12 were distributed over the values 0, 1, and 2 (with no subject scoring

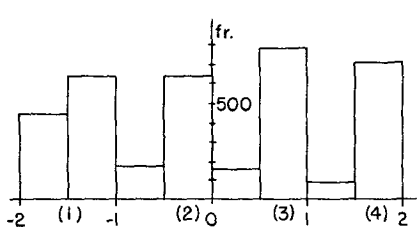


Figure 5.3.1

TABLE 5.3.4
Distribution of Third Component: Raven's Data

v_3		Group C Frequency	Group A Frequency
3.5	3.9	2	
3.0	3.4	8	4
2.5	2.9	4	
2.0	2.4	1	
1.5	1.9	34	4
1.0	1.4	45	6
.5	.9	17	1
0	.4	5	17
−.5	−.1	133	42
−1.0	−.6	114	16
−1.5	−1.1	27	3
−2.0	−1.6		
Total		390	93

3 or 4) in complete correspondence to the three suspected groups in Table 5.3.4, the explanation of the three latent classes was evident.

Similarly, in the case of Set A, the factor loadings for tests 1 and 2 were −0.725 and −0.803 respectively, while no other test had a loading of greater than ±0.201 (cf. Table 5.1.3a). Here the three apparent classes correspond precisely to 0, 1, or 2 errors in the eight items of subtests 1 and 2, so again there was no great mystery.

Discussion

In both the examples considered, there is some suggestion that the latent variate has a discrete distribution, and in both cases the basis of the discontinuity is readily understood and completely lacking in theoretical interest. This is in contrast to the constructed numerical example of a genuine latent dichotomy in sect. 4.3. It was also true in the latter case that the two values taken by the latent variate corresponded approximately to cumulated scores on the discriminating items. However, in the empirical examples, all possible total scores occurred, over the discriminating items, yielding as many latent classes. In the constructed example, there were ten discriminating items, and hence there could be eleven latent classes on this theoretically trivial basis. But the data were so constructed that only two such classes occurred.

The question for future research in this direction is this: How often

will a latent class case be discovered in empirical data, which is nontrivial theoretically? It is to be hoped that the use of the present methods of nonlinear analysis in future factor investigations will serve incidentally to answer this question.

5.5 *An Empirical Application to Latent Distance Analysis*

Introduction

In sect. 4.5, a set of relations was obtained from a Fourier analysis of a step function, which should enable one to estimate the parameters of the latent distance model, given an initial solution in the form of a latent polynomial.

A direct method for fitting a latent distance model to empirical data was first described by Lazarsfeld (1950). The currently recommended method for fitting this model directly can be found in Torgerson (1958). This makes use of the covariances of the observed variables, and also of the expected value of products of three variables. By its nature, the method imposes a rectangular distribution on the latent variate. Since the regression functions are step-functions, their form will be invariant under a change of metric of the latent continuum.

It is not yet possible to make any theoretical statements as to the relative effectiveness of the direct method in normal use and the indirect method of sect. 4.5. Some idea as to the relative efficiency of estimation of the methods could be obtained from an empirical program which would be somewhat parallel to the comparison by Hays and Borgatta (1954) of the three-parameter and the two-parameter latent distance models.

For the purposes of this section, one such comparison will be made. Hays and Borgatta (1954) give the complete score matrices, arranged in terms of response patterns and their frequencies, for ten sets of four items and five sets of five items. All fifteen would be regarded as good quasi-scales, with coefficients of reproducibility, as defined by Guttman, ranging from 0.9207 to 0.9821. Scale 10 contains 1000 subjects (with a reproducibility of 0.9403), while the next largest sample consists of 390 subjects. Accordingly, scale 10 was chosen for reanalysis by nonlinear factor analysis and the methods of sect. 4.5. Since Hays and Borgatta give no information about the items in the scale, the psychological significance of the results is unknown.

Analysis and Results

The response patterns and their frequencies are given in Table 5.5.1. The five main "scale types" are underlined. The observed covariance matrix is given in the lower triangle of Table 5.5.2, together with the factor matrix F and the latent roots C. The reproduced covariances are given in the upper triangle of the covariance matrix, with communalities in the leading diagonal.

TABLE 5.5.1

Response Pattern	Observed Frequency (0)	Component Variates			Expected Frequency
		v_1	v_2	v_3	
1111	75	2.083	1.857	.895	39.3
1110	110	1.544	−.403	−2.638	142.8
1101	14	1.257	.546	3.071	13.3
1100	141	.717	−1.714	−.463	157.4
1011	10	1.132	1.700	3.743	9.7
1010	49	.592	−.560	.210	51.1
1001	11	.306	.390	5.919	13.8
1000	161	−.234	−1.870	2.386	222.4
0111	8	.923	4.180	−1.355	6.8
0110	11	.383	1.918	−4.888	25.9
0101	8	.097	2.767	.821	3.1
0100	64	−.443	.607	−2.712	42.0
0011	3	−.028	4.022	1.494	2.8
0010	41	−.568	1.762	−2.039	29.0
0001	9	−.855	2.711	3.670	13.4
0000	285	−1.394	.451	.136	227.2

Table 5.5.1 contains also the values of the three component variates for each response pattern.

Scatter diagrams of the bivariate distributions $[v_1, v_2]$ and $[v_1, v_3]$ are given respectively in Figs. 5.5.1 and 5.5.2. Inspection of these suggests that v_2 could be represented as a quadratic function of v_1, while v_3 could be represented as a cubic, with little or no rotation. Analyses of the variances of v_2 and v_3, given in Table 5.5.3, provide confirmation of these impressions. Thus we confirm the hypothesis that the data can be accounted for by a single-factor cubic model.

The rigorous procedures for rotation given in sect. 3.2 are exceedingly arduous to carry out in three-space by hand computation. The nature of this example (see discussion section below) did not justify a refined analysis.

TABLE 5.5.2

Item	Covariance matrix				p_j	Matrix F		
1	(.1696)	.0971	.0659	.0324	.571	.359	−.188	.073
2	.0939	(.0972)	.0850	.0419	.431	.297	.014	−.094
3	.0687	.0717	(.0847)	.0575	.307	.258	.114	−.072
4	.0312	.0455	.0536	(.0807)	.138	.169	.196	.117
					C = diag	.312	.087	.033

TABLE 5.5.3

Analyses of Variance

Component Variate v_2

Source	df	SS	Ms	F
Linear	1	67.426	67.426	44.861***
Quadratic	1	347.611	347.611	231.278***
Residual	997	1,498.063	1.503	
Total	999	1,913.100		

Component Variate v_3

Source	df	SS	Ms	F
Linear	1	57.430	57.430	18.082***
Quadratic	1	33.025	33.025	10.398**
Cubic	1	218.180	218.180	68.696***
Residual	996	3,162.865	3.176	
Total	999	3,471.500		

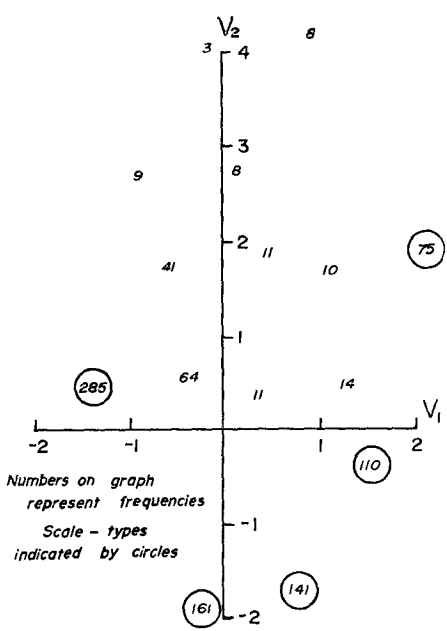


Figure 5.5.1

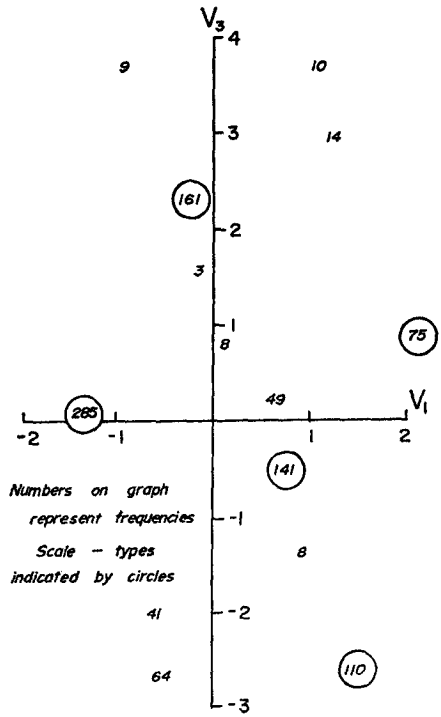


Figure 5.5.2

A comparison with the artificial example of the cubic case given in sect. 3.6 suggests that if the graphs reveal discernible quadratic and cubic parabolas, the amount of rotation required will be negligibly small, relative to other sources of error in an empirical case such as this. It was therefore assumed from inspection of the graphs that the required rotation would be near enough to zero. Table 5.5.4 gives the estimated moments of v_1, together with the moments of the true part w_1 (which we identify with the single latent variate

TABLE 5.5.4

Moments of v_1 and w_1

μ	v_1	w_1
20	1.307	1.000
30	.373	.373
40	3.224	1.096
50	2.457	11.311
60	9.947	.206

x) calculated as usual on the assumption that the disturbances are normally distributed. The moments of w_1 should give, as previously, expressions for w_2, w_3 as orthonormal polynomials in $x(= w_1)$, by (3.2.1). However, the values of the moments of w_1 so calculated are quite implausible, and in fact they yield an imaginary value for

$$k(= 1/\sqrt{\mu_4 - \mu_3^2 - 1})$$

in the expression for the quadratic component. Thus in this case the present methods for fitting polynomial regression curves have broken down completely.

In spite of the failure of the procedures described in the last paragraph, we now identify the three sets of factor loadings in Table 5.5.2 with the coefficients c_{ip} of the first, third, and second order harmonics of a step function, as indicated, and the proportion of endorsements of each item (also given in Table 5.5.2) is identified with the coefficient of the zero order harmonic. Assuming now that the latent variate has a normal distribution, we substitute these values in the expressions for c_{i0}, c_{i1}, and c_{i2} given by (4.5.1) and (4.5.2), viz.

$$c_{i0} = \beta_i + (\gamma_i - \beta_i)N(x_i)$$

$$c_{i1} = (\beta_i - \gamma_i)n(x_i)$$

$$c_{i2} = \frac{1}{\sqrt{2}}(\beta_i - \gamma_i)x_i n(x_i),$$

and thence obtain the estimates of the parameters of the latent distance model given in Table 5.5.5. The values so obtained for the "breaking points" x_i, where the regression function jumps from the value γ_i to the value β_i, correspond to a metric of the latent dimension for which x is normally distributed. To obtain results comparable with those given by Hays and Borgatta it is necessary to transform the metric so that x would have a rectangular distribution. This is achieved, in effect, by taking the transformed values for the breaking points

TABLE 5.5.5
Estimates of Model Parameters

Item	Indirect Method				Hays and Borgatta		
	β	γ	x	$x^{(n)}$	β	γ	$x^{(n)}$
1	.855	−.330	−.739	.240	.919	.081	.415
2	.822	.075	.065	.524	.846	.165	.609
3	.882	.097	.624	.732	.843	.124	.746
4	1.709	.055	1.646	.950	.961	.039	.892

$$x_i^{(n)} = N(x_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{xi} \exp\left(-\frac{z^2}{2}\right) dz$$

from a table of the normal distribution. The values of β_i, γ_i and $x_i^{(n)}$ obtained by these methods may be compared with those obtained by Hays and Borgatta using the direct method, also given in Table 5.5.5. The step-functions given by both methods are shown in Fig. 5.5.3.

A test of the relative merits of the two procedures can be obtained by calculating the expected frequencies of the response patterns from the parameters as fitted to the data. These are given in Table 5.5.1 for convenience of comparison with the observed frequencies. A figure of merit used by Hays and Borgatta to compare their three-parameter and two-parameter solutions is the sum of the absolute differences $|0 - E|$ over all response patterns. With the parameters obtained here, this figure comes to 269.6, while with the Hays and Borgatta solution it is 55.

Discussion

On this example, the indirect method of fitting a latent distance model by means of a nonlinear factor analysis yields results which are not hopelessly unrealistic; however, these results do compare unfavorably with those obtained by the direct method.

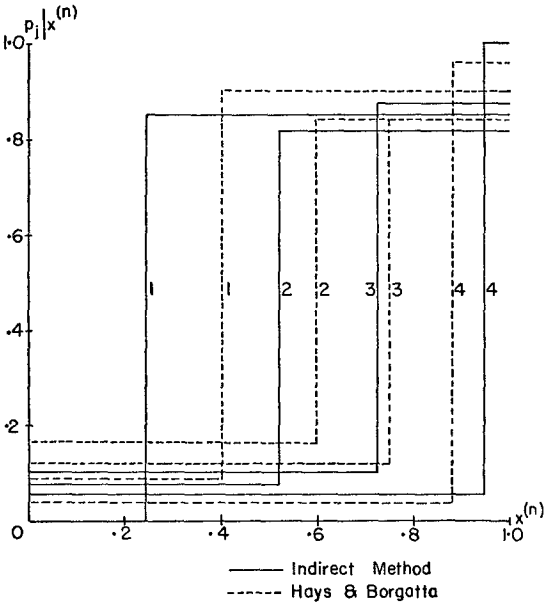


Figure 5.5.3

There are two reasons why one might not expect this example to yield closely comparable results by the two methods.

Firstly, the general theory as developed in Chap. 3 rests for its practical application on the assumption of a normal distribution for the disturbance part of the component variates, as the disturbance consists of a linear combination of the unique variations in the observed variates. This assumption can only be made with some confidence when the number of observed variables is "large." In this case we have only four observed variables, and inspection of the graphs in Figs. 5.5.1 and 5.5.2 is sufficient to dispose of any belief that the deviations in the v_1 direction from the theoretical curves could be considered normally distributed. The failure of this assumption leads in turn to absurd figures for the moments of w_1 and failure of the procedure for fitting polynomial regressions.

Secondly, in taking the step to the latent distance solution, it was suddenly assumed that the latent variate itself is normally distributed. This was done simply because, thus far, the necessary integrals in the harmonic analysis of a step function have only been written down for a normal latent distribution. However, this assumption is clearly contrary to fact! We already know at this point (cf. Table 5.5.4) that the third moment of v_1, hence of the assumed single factor x, is not zero, but 0.428, on a sample of size 1000. The assumption that the latent variate is itself normally distributed is also inconsistent with the impression given by the graphs in Figs. 5.5.1 and 5.5.2.

A third consideration with some bearing on the question of goodness of fit is the fact that the latent polynomial solution obtained above is only a first approximation, with the rotation assumed zero. It seems very unlikely that this could affect the issue, however. It was in fact because of the two major doubts just considered that a precise solution to the rotation problem did not seem worth the computational labor involved. In view of the graphs in Figs. 5.5.1 and 5.5.2 it is inconceivable that a large amount of rotation could be required for a precise solution. From experience with these methods it can be said that a small rotation would make very little difference to the solution.

It is possible that the indirect method and the direct methods for fitting a latent distance model should be considered complementary rather than, in any sense, competitive. The advantage claimed for the indirect procedure is that it can be applied in truly post facto fashion to data already analyzed in terms of the latent polynomial, if the data justify this. The possibly fatal disadvantage is that it seems unlikely to work at all well on small sets of variables. On the other hand, the computing labor in applying the direct method increases more rapidly as a function of the number of observed variables than does the labor required in the indirect method, and at the same time it is likely that the precision of the indirect method improves. (This last comment needs substantiation by an extensive program of empirical testing.)

5.8 *An Application to Learning Curves*

Introduction

In sect. 4.8 two methods were described for fitting a set of individual functional relationships using what can broadly be termed factor analytic techniques. One of these (Method I) is a direct extension of the general theory of nonlinear factor analysis, while the other (Method II) stems from the classical procedures for fitting curves by a series of orthogonal polynomials.

For a comparison of these methods in a psychological application, data were available from a variety of sources, but one set in particular seemed likely to be reasonably tractable in that the average curve for the group could be readily fitted by a quadratic function, and individual curves, on simple inspection, revealed some variation in shape, though such as should not require terms of very high order to describe.

The data chosen for analysis were collected as part of a recent unpublished study* on factors affecting the reminiscence phenomenon. For the present purpose, it will be sufficient to describe the origin of the data which was actually employed here, without a detailed account of the broader experimental context.

The task employed was the Tsai-Partington numbers test (Ammons, 1955). In the form which was constructed, this consists of a pattern of numbers (1 to 50) scattered at random on a quarto page. A specimen page is given as Fig. 5.8.1. The subject's task was to trace with pencil as rapidly as possible from one number to the next in serial order. The group whose data are to be reanalyzed here is one which received thirty-six trials on the task, with repetition of a single random pattern throughout the sequence of trials. For this group, eighteen test booklets were constructed, consisting of a cover page containing a small example of the task set, followed by thirty-six repetitions of a random pattern of numbers, and terminated by a page containing the words "STOP WORK" to avoid end-spurt. To achieve counterbalancing in the overall design of the experiment, each of the eighteen subjects in this group performed on a different random pattern, so that any individual learning function could be expected to be dependent on the particular pattern that the subject is working through, as well as on the basic learning parameters of the subject.

Instructions to the subjects emphasized the need for both speed and accuracy. Trials were timed with a stop watch. The subject's score on each trial was the highest number reached in a 30-second interval. No account was taken of inaccurate responses. These occurred very rarely in the records. Divisions between trials were indicated by the experimenter calling "stop. . .

* R. P. McDonald and A. J. Yates, "Reminiscence and perceptual search." A paper read to the 1961 annual conference of the British Psychological Society, Australian Branch.

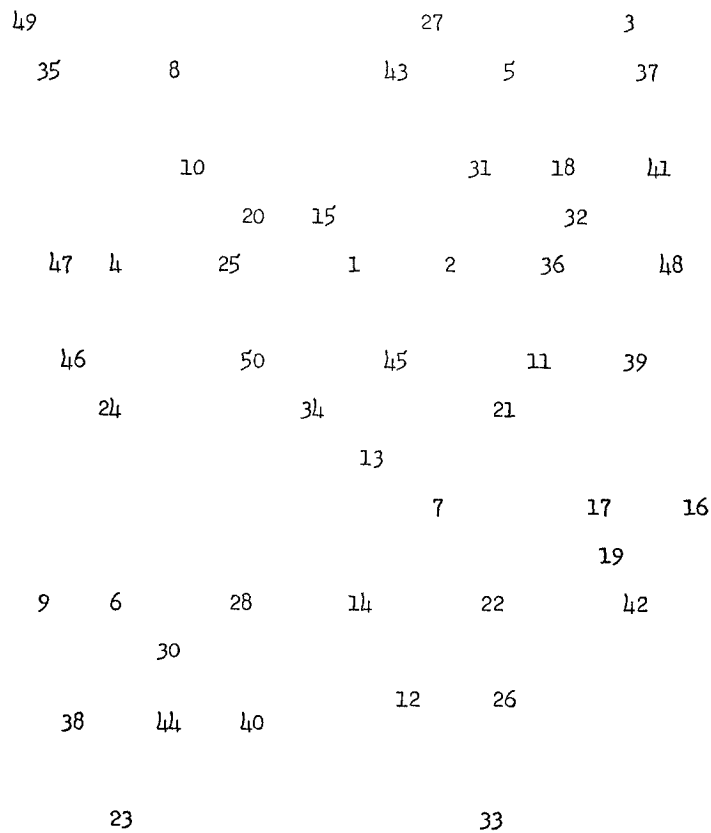


Figure 5.8.1

turn over (the page)... begin," so as to provide approximately two seconds between trials.

Analysis and Results: Method I

The score matrix Y (18 subjects $\times$ 36 trials) is given in Appendix 5.8A, together with the 18 $\times$ 18 covariance matrix (intercorrelating subjects over the 36 trials as required by the argument given in sect. 4.8). A decision was made to operate on the covariance matrix rather than on the matrix of (raw) product-moments as in the Eckart–Young procedure employed by Tucker (1960). The main motive for this was to simplify the subsequent analysis. A justification for this choice was given in sect. 2.1 where it was shown that in general, if the (raw) product-moments are analyzed, a rotation of the factor matrix exists in which the vector of means appears as a "factor." In data analyzed in this way by Tucker (1960, p. 28 and p. 35)

the first principal component (unrotated) agrees very closely with the vector of means.

In an initial principal components analysis of the covariance matrix, with variances in the leading diagonal, the five largest latent roots were 814.900, 13.111, 8.805, 7.357, 7.145.

Tucker (1960) has described a Mean Square Ratio, which is intended to provide a test of significance for the approximate rank of a score matrix, whose sampling distribution is not unlike that of Snedecor's F, but is "slightly biassed towards higher values" (Tucker, 1960, p. 17). This MSR is based, ultimately, on the theorem that the sum of the latent roots of a symmetric matrix is equal to its trace. This should enable one to compare the variance accounted for by the r'th latent root with the variance accounted for by the sum of the remaining $n - r$ latent roots. In the present notation this leads to the formula

$$(5.8.1) \qquad (MSR)_r = \frac{c_r}{(n + m - 1 - 2r)} \Bigg/ \frac{\sum_{i=r+1}^{n} c_i}{(n - r - 1)(m - r - 1)},$$

where c_i is the i'th latent root, n is the number of subjects, m the number of "trials." An adjustment has been made to the degrees of freedom, $(n + m - 1 - 2r)$ for the numerator and $(n - r - 1)(m - r - 1)$ for the denominator, to allow for the fact that we are operating here on the covariance matrix. The MSR's for the first five latent roots are, respectively, 128.584, 2.428, 1.834, 1.702, and 1.899. In view of Tucker's remarks about the sampling distribution of the MSR, and given that the total degrees of freedom in this case are comparable with Tucker's worked example, a decision was made to retain two factors, this being essentially in line with Tucker's decision. Communality estimates were therefore obtained on the hypothesis of rank two, and the covariance matrix was again subjected to a principal components analysis with these estimates in the leading diagonal. It was hoped, in any case, that this treatment of the data would be justified in the sequel. The matrix of factor loadings F and diagonal matrix of latent roots C are given in Table 5.8.1. The 36×2 matrix of component variates V was then computed. Fig. 5.8.2 gives a scatter diagram of v_2 on v_1, with the trial numbers entered on the graph at the points (v_1, v_2). Simple inspection does not strongly suggest any particular form of functional relation between them. A very faint suggestion that the early, middle, and late trials form groups which might lie on a parabola was followed up. Application of the rotation procedure for a single-factor quadratic case gave a minimum value of .795 for Φ_2, at a rotation of axes of $-17°$, but an analysis of variance of the regression of $v_2\ell_2$ on $v_1\ell_1$ gave an F of slightly less than unity for the quadratic component, hence this interpretation could not be justified.

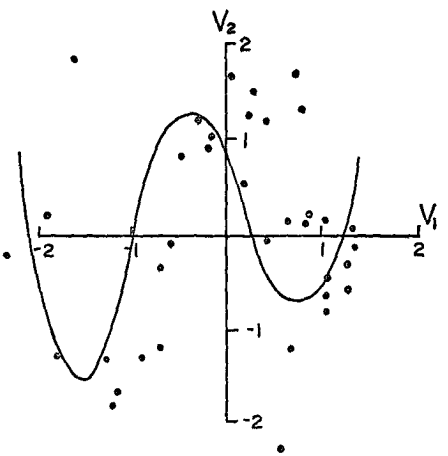
An alternative interpretation of the relation between v_1 and v_2 is sug-

TABLE 5.8.1

Factor Loading Matrix

$$F = \begin{bmatrix} 8.034 & .200 \\ 6.824 & -.824 \\ 8.880 & -.498 \\ 5.503 & .277 \\ 4.589 & -.596 \\ 9.512 & .686 \\ 8.289 & -.346 \\ 3.309 & -.974 \\ 4.271 & .554 \\ 2.781 & .456 \\ 5.435 & -.473 \\ 9.142 & -.671 \\ 6.290 & 2.506 \\ 7.323 & .077 \\ 4.904 & .211 \\ 7.298 & -.175 \\ 7.515 & -.486 \\ 5.883 & .316 \end{bmatrix}$$

$C = \text{diag}\ \{812.125 \qquad 10.826\}$



Curve : $W_2 = h_4(W_1)$

Relation between component variates

Learning curve data

Figure 5.8.2

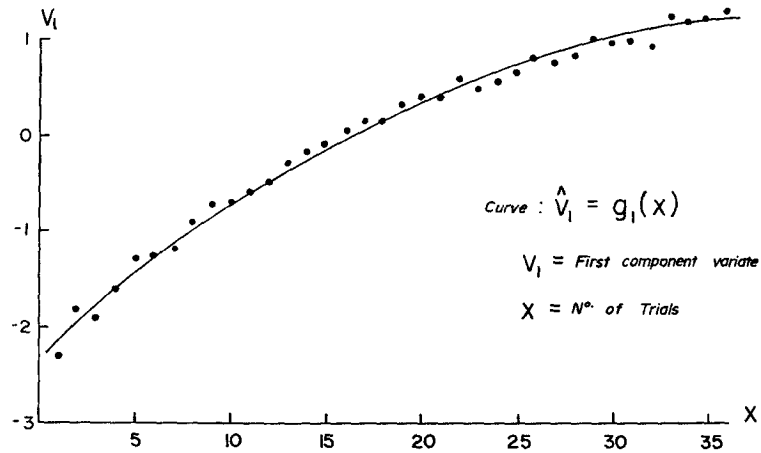


Figure 5.8.3

gested by their respective relations to the independent variable x. Graphs of these are given in Figs. 5.8.3 and 5.8.4. Regression functions for v_1 on x (number of trials) and v_2 on x were fitted to these using the tabulated orthogonal polynomials given by Pearson and Hartley (1956). Analyses of variance given in Tables 5.8.2 and 5.8.3 show that the first component variate can be

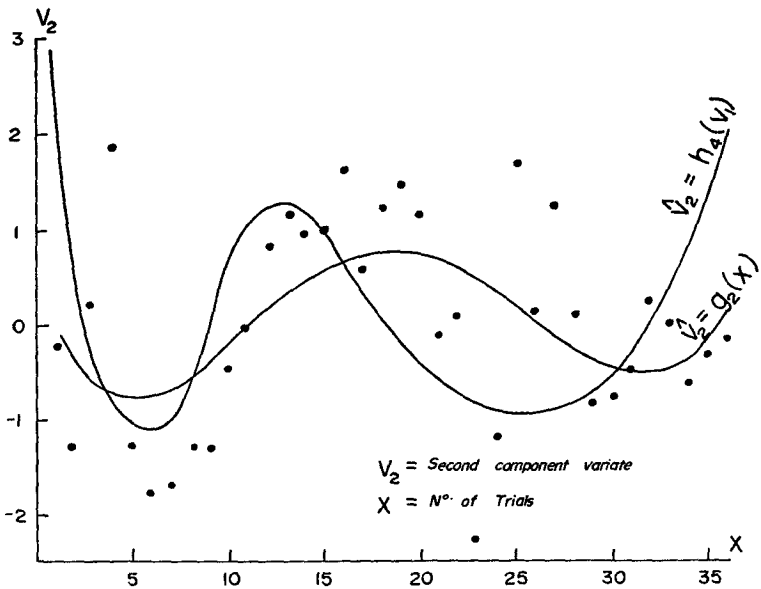


Figure 5.8.4

TABLE 5.8.2

Analysis of Variance

$\hat{v}_1 = f(x)$

Source	df	SS	Ms	F
Linear	1	34.540	34.540	4111.9***
Quad.	1	1.649	1.649	196.3***
Resid.	33	.261	.0084	
Total	35	36.450		

fitted by the quadratic function

$$\hat{v}_1 = -2.26656 + .17642x - .00222x^2 \tag{5.8.2}$$

while the second requires the quartic function

$$\hat{v}_2 = .405037 - .535978x + .076313x^2 - .003330x^3 + .000045x^4. \tag{5.8.3}$$

These functions are also plotted in Figs. 5.8.3 and 5.8.4. Since v_1 can be represented as a quadratic function in x, with very little departure from linearity, while v_2 is a quartic in x, it would be reasonable to expect that the relation between the true parts (w_1, w_2) of (v_1, v_2) could be represented by a quartic, with little or no rotation. Reexamination of Fig. 5.8.2 reveals a close resemblance to Fig. 5.8.4, with some scale distortion and misplacements in the v_1 direction. However, the analysis of variance of the regression of v_2 on v_1, given in Table 5.8.4, fails to confirm this hypothesis. In spite of this conflict of indications, this interpretation of the second factor still seemed plausible, as the misplacements of the points in Fig. 5.8.2 in the v_1

TABLE 5.8.3

Analysis of Variance

$\hat{v}_2 = f(x)$

Source	df	SS	Ms	F
Linear	1	.264	.264	<1
Quad.	1	4.430	4.430	4.152*
Cubic	1	.000	.000	<1
Quartic	1	4.539	4.539	4.254*
Resid.	31	33.067	1.067	
Total	35	42.300		

TABLE 5.8.4

Analysis of Variance

$\hat{v}_2 = f(v_1)$

Source	df	SS	Ms	F
Linear	1	.9424	.9424	1
Quad.	1	1.5832	1.5832	1.4093
Cubic	1	4.1810	4.1810	3.7217
Quartic	1	.7669	.7669	1
Resid.	31	34.8265	1.1234	
Total	35	42.3000		

direction could be sufficient to weaken the test of significance, whereas in the tests on the relations of v_1 and v_2 to x, the independent variable is not subject to error.

Retaining the hypothesis that w_2 is a quartic function in w_1, the moments of v_1 up to the eighth were estimated, and estimates of the corresponding moments of w_1 were obtained from these. These are given in Table 5.8.5. Substituting the obtained moments of w_1 in the expression for a quartic function given by (3.2.1) we obtain the function

$$h_4(w_1) = .7683 - 2.4678w_1 - 2.2947w_1^2 + 2.1526w_1^3 + 1.2609w_1^4. \tag{5.8.4}$$

This function is plotted in Fig. 5.8.2. According to the general theory, the points (v_1, v_2) are considered to be displaced by "errors" in both directions from the theoretical curve. Since the curve of $h_4(w_1)$ is obtained on the basis of the estimated distribution of v_1 only, simple inspection of the relation between the points (v_1, v_2) and the curve in Fig. 5.8.2 suggests that the interpretation made above, on the basis of the relations of v_1 and v_2 to x, remains plausible.

TABLE 5.8.5

Moments of v_1 and w_1

μ	v_1	w_1
20	1.0125	1.0000
30	−.6095	−.6095
40	2.3246	2.2490
50	−3.0686	−2.9924
60	7.8286	7.3979
70	−14.0989	−13.3034
80	32.2402	29.5764

There is a possible variant at this point on the procedures in Method I as described in sect. 4.8. We have obtained, in this case, an interpretation of the second component variate as a quartic function in the first. According to the theory of sect. 4.8, we would obtain first an expression $\hat{v}_1 = g_1(x)$, as in (5.8.2), and then obtain an expression for $\hat{v}_2$ as a function in x by substituting $g_1(x)$ for w_1 in (5.8.4). (This assumes that in the regression functions on x, which is error-free, v_1, v_2 can be identified respectively with the true components w_1, w_2.) Proceeding in this way, we obtain the function

$$\begin{aligned}(5.8.5)\qquad \tilde{v}_2 = h_4(w_1) = .7683 &- 2.4678(-2.26656 + .17642x - .00222x^2) \\ &- 2.2947(-2.26656 + .17642x - .00222x^2)^2 \\ &+ 2.1526(-2.26656 + .17642x - .00222x^2)^3 \\ &+ 1.2609(-2.26656 + .17642x - .00222x^2)^4.\end{aligned}$$

However, we already possess the expression (5.8.3) for $\hat{v}_2$ as a function in x, so that this step appears to be redundant. The function (5.8.5) is also plotted in Fig. 5.8.4. The directly obtained expression (5.8.3) necessarily gives a better fit to the data, but it is evident on inspection that the indirectly obtained expression does follow the trend of the data, in spite of the fact that it was calculated with no reference to the values of v_2.

We have, then, two solutions by Method I. In the direct solution (Method Ia, say), the *reference curves* are given by (5.8.2) and (5.8.3). In the indirect solution (Method Ib, say), the reference curves are given by (5.8.2) and (5.8.5). Individual curves are obtained by substituting the first or second set of these expressions for $\hat{v}_1$ and $\hat{v}_2$ in

$$(5.8.6)\qquad y_i = \bar{y}_i + f_{i1}v_1 + f_{i2}v_2,$$

where $\bar{y}_i$ is the mean score for the i'th subject over the 36 trials, and f_{i1}, f_{i2} are the elements of the matrix F given in Table 5.8.1. Data for a selection of subjects are plotted in Figs. 5.8.5 through 5.8.8, together with some of the curves fitted by the two variants of Method I and by Method II. In some cases, the fitted curves agree so well that it is impossible to distinguish them on the graph. To facilitate comparisons, Table 5.8.8 gives values of the fitted functions, by all methods, for selected trials.

Analysis and Results: Method II

The tabulated values of the set of orthogonal polynomials for $m = 36$, up to the fifth power, given by Pearson and Hartley (1956, p. 219) were normalized to yield the 36×5 matrix $H'(x)$ given in Appendix 5.8A. The score matrix Y was postmultiplied by $H'(x)$ according to (4.8.22) to yield the estimated matrix $\hat{K}$ given in Table 5.8.6, with an indication of the level of significance of each parameter. At this point we require to make a decision

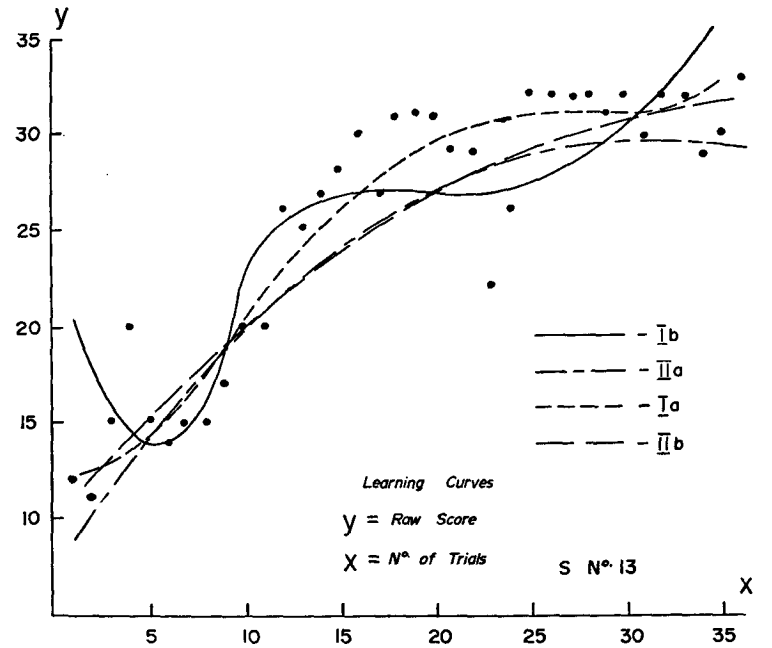


Figure 5.8.5

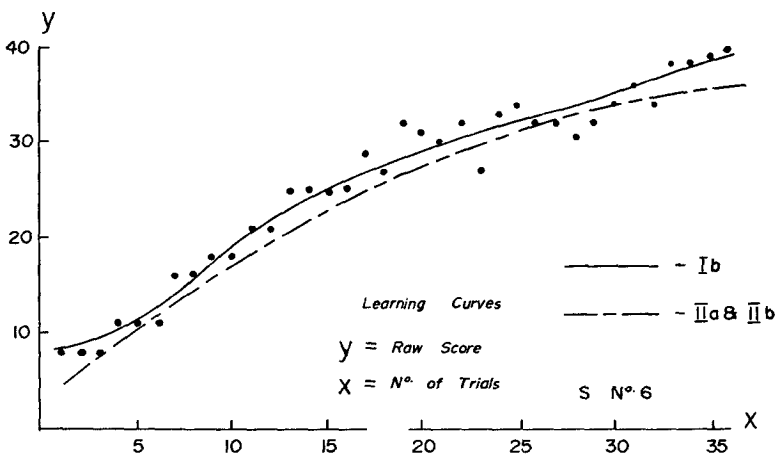


Figure 5.8.6

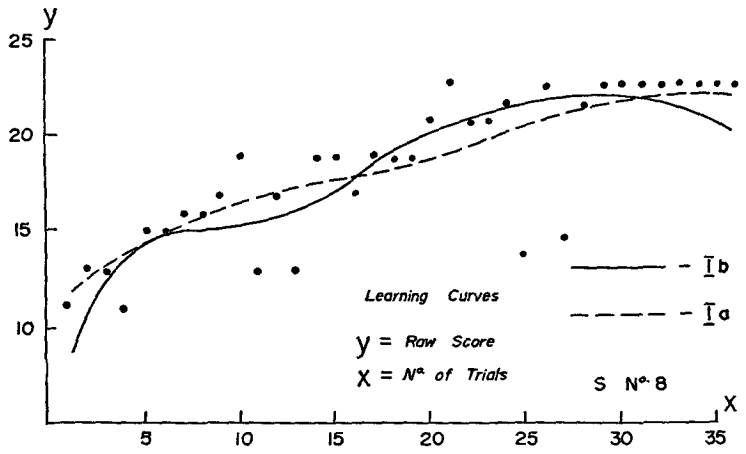


Figure 5.8.7

as to how many columns of $\hat{K}$ should be retained for subsequent analysis. Inspection of the table suggested that retention of terms in polynomials above the second degree would not be justified, as one could expect the number of technically significant results, out of eighteen, obtained in later columns, to occur by chance.*

Deleting the last three columns of $\hat{K}$ and premultiplying the 18×2 matrix ($\hat{K}$) so obtained by its transpose yields

$$\hat{K}'\hat{K} = \begin{bmatrix} 27812.80 & -6012.93 \\ -6012.93 & 1409.66 \end{bmatrix}.$$

This yields in turn the estimated principal component matrix

$$\hat{N} = \begin{bmatrix} .977262 & -.212075 \\ .212367 & .977142 \end{bmatrix}$$

and latent roots

$$C = \text{diag}\ \{29{,}117.802 \quad 104.786\}.$$

According to the argument of sect. 4.8, we hope to find (in general) that one or more of the latent roots of $\hat{K}'\hat{K}$ is zero. In a practical case, it is unlikely that this would be strictly true. While there seemed to be no theoretical justification for the introduction of an additional error term such that we would have the right to approximate $\hat{K}$ with a matrix of lower rank, the

* A comprehensive test to replace this decision-procedure would be desirable. It should be noted that the analysis by Method II was initiated at a time when the results of Method I were unknown. When both analyses were completed, it seemed best to retain the original decisions and compare the two methods as independently employed.

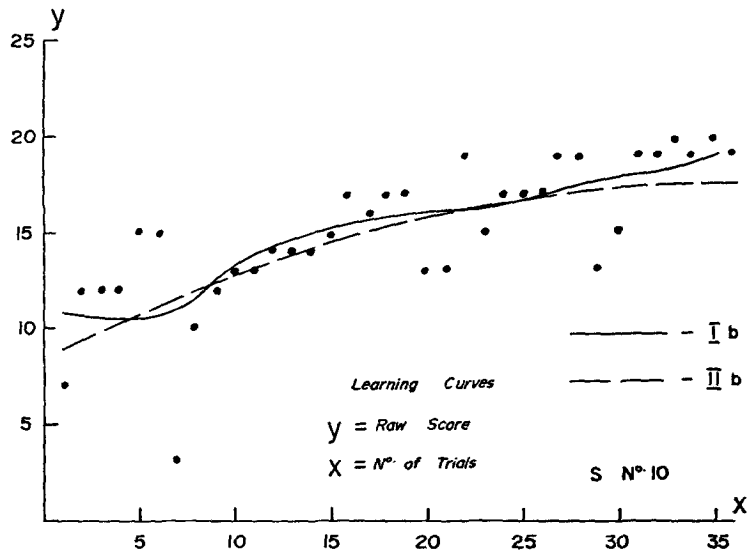


Figure 5.8.8

TABLE 5.8.6

Matrix of Regression Coefficients $\hat{K}$

	k_1	k_2	k_3	k_4	k_5
1	47.08***	−10.35***	0.99	−0.01	0.04
2	41.37***	−3.7	−4.37*	2.04	2.12
3	52.15***	−10.32***	1.00	−2.40	2.76
4	31.99***	−7.57***	0.32	0.30	−1.39
5	27.73***	−2.22	2.68	−7.40***	1.41
6	55.69***	−12.08***	4.77*	5.66**	−1.13
7	48.76***	−8.48***	6.11**	−2.49	4.77**
8	19.55***	−2.82	1.47	−0.25	0.03
9	24.41***	−7.66***	1.58	2.45	0.49
10	16.48***	−1.38	0.68	1.60	0.80
11	31.37***	−7.81***	3.42	−2.48	−1.19
12	53.17***	−12.85***	3.10	−0.46	1.64
13	35.75***	−13.92***	−0.32	5.05	−2.97
14	43.29***	−8.33***	0.17	−1.29	1.05
15	28.95***	−5.68***	1.95	−0.08	−4.61**
16	42.32***	−10.04***	2.67	1.70	−2.24
17	43.59***	−11.46***	1.95	−1.39	0.47
18	33.94***	−8.94***	5.26**	1.08	1.89

practical implication of the theory is that we should attempt to do just this. That is, we hope to obtain an adequate description, though *necessarily* less precise, of the individual learning curves, by deleting rows of $\hat{N}$ which correspond to "negligibly small" latent roots.

In this case we choose to delete the second row of $\hat{N}$, regarding the second latent root (104.786) as "negligibly small," giving the reduced matrices

$$\tilde{N} = [.977262 \quad -.212075]$$

and

$$\tilde{C} = 29{,}117.802.$$

Postmultiplying $\tilde{K}$ by $\tilde{N}$ yields the column vector $\tilde{M}\tilde{C}^{1/2}$ given in Table 5.8.7. The success of the approximation can be judged from the product matrix $\tilde{M}\tilde{C}^{1/2}\tilde{N}$ given also in Table 5.8.7.

The general expressions for the orthonormal polynomials in

$$H(x) = [h_1(x)h_2(x)]$$

are

$$h_1(x) = 0.01604x \quad -0.29683$$

and

$$h_2(x) = 0.00173x^2 \quad -0.06396x + 0.52946.$$

TABLE 5.8.7

$\tilde{M}\tilde{C}^{1/2}$	$\tilde{M}\tilde{C}^{1/2}\tilde{N}$	
48.191	47.10	−10.22
41.207	40.27	−8.74
53.138	51.93	−11.27
32.859	32.11	−6.97
27.563	26.94	−5.85
56.970	55.67	−12.08
49.436	48.31	−10.48
19.698	19.25	−4.18
25.472	24.89	−5.40
16.394	16.02	−3.48
32.304	31.57	−6.85
54.671	53.43	−11.59
37.879	37.02	−8.03
44.060	43.06	−9.34
29.488	28.82	−6.25
43.475	42.49	−9.22
45.017	43.99	−9.55
35.055	34.26	−7.43

These are obtained by normalizing the expressions given by Pearson and Hartley (1956, p. 91). Premultiplying this expression for $H(x)$ by $\tilde{N}$ gives the *reference curve*

$$\tilde{T}(x) = \tilde{N}H(x) = -0.402366 + 0.029240x - 0.000367x^2.$$

Individual learning curves are then given by $\tilde{M}\tilde{C}^{1/2}\tilde{T}(x)$. A selection of these curves is included in Figs. 5.8.5 through 5.8.8, together with the curves given by

$$\tilde{Y} = \tilde{K}H(x)$$

and the curves given by the two variants of Method I. Where the agreement between the curves is too close to represent them on the same graph, comparisons can be made by an examination of selected values of the fitted functions given in Table 5.8.8.

Discussion

The two variants on Method I, the basic procedure for fitting curves by orthogonal polynomials, and the "condensation" performed on these by Method II yield altogether four sets of functions exemplified in Figs. 5.8.5 through 5.8.8 and Table 5.8.8. All four would appear to follow the trend of the data reasonably well, and in this case there is little ground for choosing between them when we balance goodness of fit against economy of description.

On the whole, the advantages seem to lie with Method II. Computing procedures are very much easier with this method, and at no stage in the procedure are there any basic decisions to be made as to the nature of the model required. In Method I, at least in one variant, it may be necessary in more complicated cases to find a rotation, in a space of more than two dimensions, for the component variates and to identify the nonlinear factor model which best accounts for their relationships. This will always be less straightforward than the procedures in Method II.

A disadvantage with both methods is the restriction to a description in terms of polynomials. In this kind of application there are often good theoretical grounds for prescribing a "law" of a different mathematical type such as the exponential growth function. It may prove possible, as in the treatment of the normal ogive model and the latent distance model in sects. 4.4 and 4.5, to obtain the parameters of such prescribed functions from an initial solution in terms of polynomials.

It should also be noted that if the *reference curves* are to be thought of as having some theoretical significance in their own right, it may seem unsatisfactory to require that these functions be mutually orthogonal. In the present example, v_1 is a monotonic function in x ($x = 1, \cdots, 36$) and could therefore qualify as a learning curve. Clearly, v_2 cannot be thought of as a learning function. In general, oblique solutions will be required if one is seeking for theoretical significance in the reference curves.

TABLE 5.8.8

Representative Fitted Curves

	Ia	*Ib*	*IIa*	*IIb*
Trial		Subject No. 13		
1	12.05	20.67	8.82	11.21
5	14.41	13.68	14.10	15.31
10	20.30	23.16	19.62	19.80
15	26.06	27.00	23.94	23.59
20	29.68	26.70	27.06	26.68
25	30.83	27.58	28.98	29.07
30	30.88	30.48	29.70	30.76
35	32.89	36.54	29.22	31.75
		Subject No. 6		
1	5.73	8.09	4.39	4.39
5	11.45	11.25	10.55	10.55
10	18.64	19.43	17.30	17.31
15	24.93	25.18	23.01	23.01
20	29.78	28.95	27.66	27.67
25	33.09	32.17	31.27	31.27
30	35.25	35.08	33.82	33.83
35	37.11	37.99	35.33	35.33
		Subject No. 8		
1	11.66	8.32	11.72	11.17
5	14.50	14.78	13.58	13.31
10	16.33	15.22	15.67	15.66
15	17.56	17.20	17.52	17.67
20	18.99	20.16	19.11	19.32
25	20.72	22.01	20.46	20.63
30	22.22	22.45	21.55	22.19
35	22.30	21.03	22.40	22.19
		Subject No. 10		
1	9.18	10.75	9.76	8.91
5	10.68	10.55	11.12	10.68
10	12.94	13.45	12.74	12.62
15	14.95	15.13	14.25	14.27
20	16.66	15.89	15.67	15.61
25	17.83	16.67	16.98	16.66
30	18.89	17.64	18.20	17.40
35	20.52	19.01	19.31	17.85

$$Ia - \hat{y} = \bar{y} + f_{j1}\hat{v}_1 + f_{j2}\hat{v}_2$$
$$Ib - \hat{y} = \bar{y} + f_{j1}\hat{v}_1 + f_{j2}h_4(\hat{v}_1)$$
$$IIa - \hat{y} = \bar{y} + k_1h_1(x) + k_2h_2(x)$$
$$IIb - \hat{y} = \bar{y} + \tilde{M}\tilde{C}^{1/2}\tilde{C}(x)$$

CHAPTER 6

SUMMARY AND CONCLUSIONS

The theory of nonlinear factor analysis presented here was based initially on the demonstration that the fundamental equations usually associated with linear factor analysis still hold under much more general conditions (sect. 2.1). A restatement in general terms of the principles of Latent Structure Analysis (sect. 2.3) revealed that linear and nonlinear factor analysis can be regarded as particular cases in LSA, as well as the recognized models. It also became apparent (sect. 2.4) that the distinction between "quantitative" or multivalued observations and "qualitative" or category observations is not of fundamental importance in distinguishing factor models.

Following the examination of an analogy (sect. 2.5) with the procedures for fitting single or multiple curvilinear regressions by orthogonal polynomials, the basic relations for a method of nonlinear factor analysis were introduced (sect. 3.1). The basic principle is that corresponding to any particular nonlinear model there should be a restricting of the distribution of component variates to a curvilinear manifold spanning the space defined by an orthogonal "linear" factor analysis. A complete, formal treatment was given (sects. 3.2 and 3.3) for the single-factor and multiple-factor polynomial models, together with an informal introduction (sect. 3.4) to the class of cases where terms in products of factor scores (latent variates) appear in the specification equation of the model. Essentially, this treatment involves estimating the moments of the "true" parts of the observed component variates, the latter being thought of as "disturbed" by unique variations from their true positions on the curvilinear manifold prescribed by the assumed model. Using these estimated moments of the "true" parts, a rotation can be found which simplifies the description of the curvilinear manifold. This in turn allows one to apply analyses of variance to the relations between the rotated component variates to test the significance of the curvilinear relations implied by the model assumed.

Certain specializations on the general theory were then presented. It was shown how a "difficulty factor" might be expected to arise from the variations in curvature of the regression functions from "easy" to "difficult" tests or items (sect. 4.1). The latent polynomial model discussed but not analyzed by Lazarsfeld was shown to be a special case of the single-factor polynomial treated earlier, hence readily analyzed by the present methods (sect. 4.2). The possibility of detecting a discrete distribution of a latent

variate was demonstrated (sect. 4.3). This procedure has the advantage over conventional methods of latent class analysis that one would not impose a latent class solution on data which "really" possess latent continuity. Methods for deriving the parameters of the normal ogive and latent distance models from an initial solution in terms of polynomials were given (sects. 4.4 and 4.5). These again have the advantage over conventional methods of avoiding the imposition of such models on data to which they are inapplicable, though in these two cases users of the conventional methods would usually avoid this if they were not applying the methods blindly. The "perfect scale," or "ideal answer pattern" of Walker was treated as a limiting case of the latent distance and normal ogive models (sect. 4.6). In contrast to Guttman's psychological interpretation of the principal components of the perfect scale, these were treated as the harmonics of a step function. This last point was developed further (sect. 4.7) into a demonstration that in addition to the factor model with which Guttman prefers to account for the *simplex property*, the step-function regression model (single-factor) is an alternative to be taken seriously. The final specialization on the general theory concerned an extension of Tucker's work on the fitting of a set of functional relationships by linear factor analysis (sect. 4.8). One method was described which stems directly from the general theory of nonlinear factor analysis. A second method was also introduced which involves the factor analysis (in a broad sense) of a matrix of individual parameters of the regression functions obtained by the classical methods of curve fitting by orthogonal polynomials.

Empirical material was sought that would serve both to illustrate the general principles of nonlinear factor analysis and to exemplify some at least of the specializations on this theory. A factor analysis of subtests of the Raven PM (1947) (sect. 5.1) served to verify the interpretation of a "difficulty factor" as due to differential curvature. In a corresponding analysis based on a second sample, the "difficulty factor" failed to emerge, but a second factor due to differential curvature was still obtained. This example serves to underline a disturbing implication of the whole of the present theory. Hundreds of studies using factor analysis have appeared by now in the psychological literature. It seems probable that a number of the factors that have been found, interpreted, and neatly labelled as a dimension of ability or personality, would prove on reanalysis to be components due to nonlinear regressions of the tests on the "real" or "basic" factors. This is most likely in the case of those investigators who enthusiastically extract a rather large number of factors from a rather small matrix. It is hardly to be expected that all existing studies be repeated, using the methods advocated here, but there would be a strong case for repetition of a number of major studies or reanalysis where full records of the initial score matrices have been kept.

A nonlinear factor analysis was carried out (sect. 5.2) on two quasi-scales put together in a single matrix. The covariance matrix was readily accounted for, initially, by three factors, and the ultimate solution reduced to a set of quadratic functions in one latent variate, together with linear functions in a second. There are two general implications which are illustrated by this study. Firstly, in a nonlinear model, rotation to achieve "simple structure" in the sense of Thurstone is no longer strictly possible, in general, and rotation to achieve "positive manifold" may be obtained at the price of considerable complications in the description of the regression surfaces. Secondly, one recognizes that a quasi-scale will require regression functions, for the items on the factor, lying somewhere between the two extremes of a straight line and a step function. This example suggests that quasi-scales with quite high coefficients of reproducibility do not need a very large departure from linearity in the regression functions to account for them. The highest degree of the required polynomial (in this case, two) is much nearer to unity than might have been expected.

Investigations (sect. 5.3) in the direction of latent class models (in the sense of a discrete distribution of the latent variate) revealed one negative instance and two trivial ones. The positive implication of these cases is that conventional latent class analysis should be supplemented (or replaced) by a nonlinear factor analysis if one does not wish to impose a latent class interpretation on data irrespective of the actual nature of the latent distribution.

A reanalysis (sect. 5.5) of data treated by Hays and Borgatta in terms of the latent distance model failed quite badly. Firstly, it gave nonsense results (nonreal coefficients) in the attempt to fit a latent polynomial to the data. Secondly, the latent distance parameters estimated by the present methods gave a considerably worse fit to the observed response frequencies than the parameters estimated by Hays and Borgatta using the direct methods of analysis for this model. These failures were understandable, since it was possible to detect the fact that two of the assumptions made in the theory of the present methods had been quite grossly violated. It is still possible that these methods would work, and reveal some advantages, in applications to data based on reasonably large sets of items.

The final empirical illustration (sect. 5.8) was an application of the two methods described earlier to the fitting of individual learning curves. Both methods "work," in the sense that they supply an economical description of the individual curves, with reasonable fit. On the whole, the advantage seems to lie with the method based on classical curve-fitting procedures, rather than with the method based on nonlinear factor analysis. The "classical" procedure involves easier computations, and at all stages in the procedure it is somewhat less "mysterious."

At this point, we turn to certain general issues which deserve review

or closer examination. These include a review of the theoretical and practical restrictions on the present approach to nonlinear factor analysis and some remarks on the questions of factor metric, the invariance problem, and rotation.

Theoretically, the restrictive assumptions involved in the methods given here are not in fact very restrictive. The assumption that the disturbance parts of the component variates have a normal joint density function can be justified for a large set of observed variates on the basis of the central limit theorem. Failure of this assumption can be detectable, as in the example of sect. 5.5. The assumption that the regression functions can be adequately described by polynomials seems reasonable, in that the description of any curve can be approximated as closely as we please by a polynomial of sufficiently high degree. The assumption that, in any particular case, linear terms in the latent variates are present, could fail in some cases (one such case was discussed in sect. 3.4). This assumption is in part related to the question of factor metric, which will be briefly treated below.

On the other hand, although in principle any curve can be described by polynomials of sufficiently high degree, in practice one could not expect to detect the higher order components (say, above the third degree) from the relations between the component variates, unless the communalities are very much higher than are usually obtained in psychological work and the disturbances due to unique variations are correspondingly small.

It may have seemed rather curious that no assumptions were introduced in Chap. 3 as to the distribution of the latent variates on the latent continuum, nor were there any assumptions as to the metric of the latent continuum (the *factor metric*). An origin and a unit were defined for each latent variate or factor by setting its mean to zero and its variance to unity, but nothing was required beyond this. On the other hand, the assumption *was* made, not in complete innocence, that linear terms in the latent variates are present in the model. This assumption and the procedures following serve to define the factor metric, and in terms of this metric the distribution of the latent variates emerges "empirically." After the analysis is completed, it would always be possible to subject the linear terms in the fitted model to any monotonic transformation. With corresponding transformations applied to the higher order functions in the latent variates, we would alter both the factor metric (and *distribution*) and the regression functions, without affecting the relations between the observed variates for which they jointly account. An example of this is the transformation applied to the "breaking-point" parameters of the latent distance model in sect. 5.8, such that the latent distribution becomes rectangular.

One of the general problems in factor analytic theory concerns the invariance or otherwise of results obtained under selection of subjects, tests, or items. Rotation of the matrix of factor loadings to "simple structure,"

orthogonal or oblique, has been considered to have advantages in the invariance of the overall pattern of zero versus nonzero loadings from one study to another. (This is essentially due to the simple fact that zero multiplied by any coefficient is still zero.) We have already seen that the concept of simple structure may be difficult or impossible to redefine in the context of nonlinear factor analysis. The whole question of factorial invariance will need to be reappraised in this context. A small beginning has been made on this, in the form of the treatment developed ad hoc for finding an optimum point on the Group C and Group A latent dimensions (sect. 5.1) for linking up the two sets of regression functions. A problem for the future will be the development of a more general procedure on these lines or perhaps by other means.

We turn now to a general issue in factor analysis which has aroused a good deal of controversy. This is the matter of oblique, as against orthogonal, solutions. Thurstone (1947) adopted the position that factors, conceived of as fundamental dimensions of ability, personality, and the like, need not be thought of as orthogonal. After all, such fundamental and *observable* dimensions of physique as height and weight are demonstrably correlated. There is a good deal of impressive argument, by analogy with observable dimensions of physical entities, to justify this view. For our purposes, however, it proved necessary to stipulate that the latent variates or factors be not merely orthogonal, but completely mutually independent in the probability sense. If this stipulation is not made, there can be no distinction between linear and nonlinear factor analysis. A fortiori, we cannot allow the factors to be correlated. Presumably, it is not possible to prove by a process of definition that the advocates of oblique solutions are wrong. Just as it is possible, after the main analysis, to transform the metric of the factor space to one which may yield desirable properties in the solution, so it is not inconceivable that one could derive a "desirable" oblique solution from an initial solution in terms of statistically independent factors. This will almost certainly prove a more complex task than the corresponding one in linear factor analysis.

The final points to emphasize are these: The present treatment of nonlinear factor analysis has its basis in a mathematical model of great generality. This leads to a single basic procedure for the dimensional analysis of data, with a rational choice of the specific model which best accounts for it. Although a good deal of further development is required, the direction that this work should take is already clear. At its present stage the model yields both a unifying theory of factor models and a method of analysis which can already be applied to empirical data to yield worthwhile results.

BIBLIOGRAPHY

Ahmavaara, Y. (1957) *On the unified factor theory of mind.* Helsinki: Suomalaisen Tiedeakatemia.

Albert, A. A. (1944) The minimum rank of a correlation matrix. *Proc. Nat. Acad. Sci., 30,* 144–146.

Ammons, C. H. (1955) Task for the study of perceptual learning and performance variables. *Percept. mot. Skills, 5,* 11–14.

Anderson, T. W. (1954) On the estimation of parameters in latent structure analysis. *Psychometrika, 19,* 1–10.

Anderson, T. W. (1959) Some scaling models and estimation procedures in the latent class model. In U. Grenander (Ed.), *Probability and statistics: The Harold Cramer volume.* New York: Wiley.

Anderson, T. W., and Rubin, H. (1956) Statistical inference in factor analysis. *Proceedings of the third Berkeley symposium on mathematical statistics and probability, 5,* 111–150.

Bargmann, R. (1960) (Review) Ahmavaara, On the unified factor theory of mind. *Psychometrika, 25,* 105–109.

Bartlett, M. S. (1962) Essays on probability and statistics. New York: Wiley.

Borgatta, E. F. (1958) On analyzing correlation matrices: Some new emphases. *Publ. opin. Quart., 22,* 516–528.

Burt, C. (1937) Correlations between persons. *Brit. J. Psychol., 28,* 59–96.

Burt, C. (1950) The factorial analysis of qualitative data. *Brit. J. Psychol., Statist. Sect., 3,* 166–185.

Burt, C. (1952) Tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sect., 5,* 109–133.

Burt, C. (1953) Scale analysis and factor analysis: Comments on Dr. Guttman's paper. *Brit. J. statist. Psychol., 6,* 5–24.

Carroll, J. B. (1945) The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika, 10,* 1–20.

Cattell, R. B. (1952) *Factor analysis.* New York: Harper.

Cattell, R. B. (1960) Evaluating interaction and non-linear relations by factor analysis. *Psychol. Rep., 7,* 69–70.

Cliff, N. (1962) Analytic rotation to a functional relationship. *Psychometrika, 27,* 283–296.

Comrey, A. L., and Levonian, E. (1958) A comparison of three point coefficients in factor analyses of MMPI items. *Educ. psychol. Measmt, 18,* 739–755.

Coombs, C. H. (1960) A theory of data. *Psychol. Rev., 67,* 143–159.

Coombs, C. H., and Kao, R. C. (1954) On the multidimensional analysis of monotonic single stimuli data. In C. H. Coombs, R. M. Thrall, and R. L. Davis (Eds), *Decision processes.* New York: Wiley.

Coombs, C. H., and Kao, R. C. (1955) Non-metric factor analysis. Engng. Res. Bull. No. 38. Ann Arbor: Univer. Michigan Press.

Dingman, H. F. (1958) The relation between coefficients of correlation and difficulty factors. *Brit. J. statist. Psychol., 11,* 13–18.

DuBois, P. H. (1960) An analysis of Guttman's simplex. *Psychometrika, 25,* 173–182.

Eckart, C., and Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika, 1,* 211–218.

Eysenck, H. J. (1958) The nature of anxiety and the factorial method. *Psychol. Rep., 4,* 453–454.

Feller, W. (1957) *An introduction to probability theory and its applications,* Vol. 1. (2nd ed.) New York: Wiley.

Ferguson, G. A. (1941) The factorial interpretation of test difficulty. *Psychometrika, 6,* 323–329.

Gabriel, K. R. (1954) The simplex structure of the PM test. *Brit. J. statist. Psychol., 7,* 9–14.

Gibson, W. A. (1955) An extension of Anderson's solution for the latent structure equations. *Psychometrika, 20,* 69–74.

Gibson, W. A. (1959) Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika, 24,* 229–252.

Gibson, W. A. (1960) Nonlinear factors in two dimensions. *Psychometrika, 25,* 381–392.

Goodman, L. A. (1959) Simple statistical methods for scalogram analysis. *Psychometrika, 24,* 29–43.

Gourlay, N. (1951) Difficulty factors arising from the use of tetrachoric correlations in factor analysis. *Brit. J. Psychol., Statist. Sect., 4,* 65–73.

Green, B. F. (1951) A general solution for the latent class model of latent structure analysis. *Psychometrika, 16,* 151–166.

Green, B. F. (1952) Latent structure analysis and its relation to factor analysis. *J. Amer. statist. Assoc., 47,* 71–76.

Guilford, J. P. (1936) *Psychometric methods.* New York: McGraw-Hill.

Guilford, J. P. (1941) The difficulty of a test and its factor composition. *Psychometrika, 6,* 67–77.

Guttman, L. (1941) In P. Horst (Ed.), *The prediction of personal adjustment.* New York: Soc. Sci. Res. Council.

Guttman, L. (1950) The principal components of scale analysis. In S. S. Stouffer et al. (Eds), *Measurement and prediction.* Princeton, N. J.: Princeton Univer. Press. Pp. 312–361.

Guttman, L. (1953) A note on Sir Cyril Burt's "Factorial analysis of qualitative data." *Brit. J. statist. Psychol., 6,* 1–4.

Guttman, L. (1954a) Some necessary conditions for common-factor analysis. *Psychometrika, 19,* 149–162.

Guttman, L. (1954b) A new approach to factor analysis: the radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press. Ch. 6.

Guttman, L. (1954c) The principal components of scalable attitudes. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press. Ch. 5.

Guttman, L. (1955a) The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *Brit. J. statist. Psychol., 8,* 65–82.

Guttman, L. (1955b) A generalized simplex for factor analysis. *Psychometrika, 20,* 173–192.

Guttman, L. (1956) "Best possible" systematic estimates of communalities. *Psychometrika, 21,* 273–285.

Guttman, L. (1957a) Simple proofs of relations between the communality problem and multiple correlation. *Psychometrika, 22,* 147–158.

Guttman, L. (1957b) Empirical verification of the radex structure of mental abilities and personality traits. *Educ. psychol. Measmt, 17,* 391–407.

Harman, H. (1960) *Modern factor analysis.* Chicago: Univer. Chicago Press.

Hays, D. G., and Borgatta, E. F. (1954) An empirical comparison of restricted and general latent distance analysis. *Psychometrika, 19,* 271–280.

Henrysson, S. (1957) *Applicability of factor analysis in the behavioral sciences.* Stockholm: Almqvist and Wiksell.

Holzinger, K. J., and Harman, H. H. (1941) *Factor analysis.* Chicago: Univer. Chicago Press.

Humphreys, L. G. (1960) Investigations of the simplex. *Psychometrika, 25,* 313–324.

Kaiser, H. F. (1961) A note on Guttman's lower bound for the number of common factors. *Brit. J. statist. Psychol., 14,* 1–2.

Kendall, M. G. (1951) *The advanced theory of statistics.* Vol. 2. London: Griffin.

Lawley, D. N. (1940) The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinburgh, 60,* 64–82.

Lazarsfeld, P. F. (1950) Ch. 10, The logical and mathematical foundation of latent structure analysis, and Ch. 11, The interpretation and computation of some latent structures. In S. S. Stouffer et al. (Eds), *Measurement and prediction.* Princeton, N. J.: Princeton Univer. Press. Pp. 362–412; 413–472.

Lazarsfeld, P. F. (1954) A conceptual introduction to latent structure analysis. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press. Ch. 7.

Lazarsfeld, P. F. (1959) Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science,* Vol. 3. New York: McGraw-Hill.

Lazarsfeld, P. F. (1960) Latent structure analysis and test theory. In H. Gulliksen and S. Messick (Eds), *Psychological scaling: Theory and applications.* New York: Wiley. Pp. 83–95.

Ledermann, W. (1937) On the rank of the reduced correlation matrix in multiple factor analysis. *Psychometrika, 2,* 85–93.

Loevinger, Jane. (1947) A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr., 61,* No. 4 (Whole No. 285).

Loevinger, Jane. (1957) Objective tests as instruments of psychological theory. *Psychol. Rep., 3,* 635–694.

Lord, F. M. (1953) An application of confidence intervals and of maximum likelihood to the examination of an examinee's ability. *Psychometrika, 18,* 57–76.

Lord, F. M. (1958) Some relations between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika, 23,* 291–296.

McDonald, R. P. (1962a) A note on the derivation of the general latent class model. *Psychometrika, 27,* 203–206.

McDonald, R. P. (1962b) A general approach to nonlinear factor analysis. *Psychometrika, 27,* 397–415.

McDonald, R. P. (1965a) Numerical methods for polynomial models in nonlinear factor analysis. *Res. Bull.* Princeton, N. J.: Educational Testing Service. RB-65-32.

McDonald, R. P. (1965b) Some IBM 7090–94 programs for nonlinear factor analysis. *Res. Mem.* Princeton, N. J.: Educational Testing Service. RM-65-11.

McDonald, R. P. (*in press*) Factor interaction in nonlinear factor analysis. *Res. Bull.* Princeton, N. J.: Educational Testing Service.

McHugh, R. B. (1956) Efficient estimation and local identification in latent class analysis. *Psychometrika, 21,* 331–348.

Madansky, A. (1960) Determinantal methods in latent class analysis. *Psychometrika, 25,* 183–196.

Maxwell, A. E. (1957) Contour analysis. *Educ. psychol. Measmt, 17,* 347–360.

Maxwell, A. E. (1959a) A statistical approach to scalogram analysis. *Educ. psychol. Measmt, 19,* 337–350.

Maxwell, A. E. (1959b) Statistical methods in factor analysis. *Psychol. Bull., 56,* 228–235.

Miller, C. R., Sabagh, G., and Dingman, H. F. (1962) Latent class analysis and differential mortality. *J. Amer. statist. Assoc., 57,* 430–438.

Mosteller, F. (1958) The mystery of the missing corpus. *Psychometrika, 23,* 279–290.

Pearson, E. S., and Hartley, H. O. (1956) *Biometrika tables for statisticians*, Vol. 1. Cambridge: Cambridge Univer. Press.

Rao, C. R. (1955) Estimation and tests of significance in factor analysis. *Psychometrika, 20,* 93–112.

Richardson, M. W. (1936) Relation between the difficulty and the differential validity of a test. *Psychometrika, 1,* 33–49.

Saunders, D. R. (1956) Moderator variables in prediction. *Educ. psychol. Measmt, 16,* 209–222.

Schuessler, K. F. (1961) A note on the statistical significance of a scalogram. *Sociometry, 24,* 312–318.

Schutz, W. (1958) *FIRO—A three-dimensional theory of interpersonal relations.* Holt, Rinehart & Winston.

Suchman, E. A. (1950) The intensity component in attitude and opinion research. In S. S. Stouffer et al. (Eds), *Measurement and prediction.* Princeton, N. J.: Princeton Univer. Press. Pp. 213–276.

Szëgo, G. (1959) Orthogonal polynomials. *Amer. math. Soc. Coll. Publ., 23.*

Thomson, G. H. (1950) *The factorial analysis of human ability.* (4th ed.) London: Univer. London Press.

Thurstone, L. L. (1947) Multiple factor analysis. Chicago: Univer. Chicago Press.

Torgerson, W. S. (1958) *Theory and methods of scaling.* New York: Wiley.

Tucker, L. R. (1958) Determination of the parameters of a functional relation by factor analysis. *Psychometrika, 23,* 19–24.

Tucker, L. R. (1960) Determination of generalized learning curves by factor analysis. ONR Technical Report. Princeton, N. J.: Educational Testing Service.

Walker, D. A. (1931) Answer pattern and score scatter in tests and examinations. *Brit. J. Psychol., 22,* 73–86.

Walker, D. A. (1936) Answer pattern and score scatter in tests and examinations. *Brit. J. Psychol., 26,* 301–308.

Walker, D. A. (1940) Answer pattern and score scatter in tests and examinations. *Brit. J. Psychol., 30,* 248–260.

Wherry, R. J., and Gaylord, R. H. (1944) Factor pattern of test items and tests as a function of the correlation coefficient: Content, difficulty and constant error factors. *Psychometrika, 9,* 237–244.

Wrigley, C. F. (1957) The distinction between common and specific variance in factor theory. *Brit. J. statist. Psychol., 10,* 81–98.

Wrigley, C. F. (1958) Objectivity in factor analysis. *Educ. psychol. Measmt, 18,* 463–476.

Yates, A. J. (1961) Item analysis of progressive matrices (1947). *Brit. J. educ. Psychol., 31,* 152–157.

APPENDIX 3.4

Some Generalizations on the Conjunctive/Disjunctive Models

In the course of developing factor models free of strong assumptions with respect to the metric of the factor space, Coombs (cf. Torgerson, 1958) has described two formally identical models, the *conjunctive* and *disjunctive* models, which have a number of interesting properties. In the *conjunctive* model, an individual passes an item if his measure on each relevant attribute is not less than the difficulty level of the item. In the *disjunctive* model, an individual passes an item if he possesses a certain minimum value on any one relevant dimension. Because of the formal identity between the two models, discussion will be restricted to the disjunctive model in the following. Existing methods of data analysis in terms of the disjunctive model make no allowance for error. Thus the usefulness of the model in practical applications is somewhat restricted.

In this section we consider some generalizations embodying the essential logic of the model, some of which are open to treatment by nonlinear factor analysis. Discussion will be restricted to cases involving two factors only. Further generalization to higher spaces is reasonably obvious.

It is convenient to describe the disjunctive model in what may appear a rather inelegant fashion. We specify the position of any individual in a two-dimensional factor space by coordinates (x, y), these being statistically independent latent variates. The item j can be described by two parameters (x_i, y_i). These may be thought of as the coordinates of the item, if we treat it as a point in the same space. In contrast to Coombs' treatment, the parameters and latent variates may here assume both positive and negative values.

Separate statements are required for individual points lying in the half plane described by

$$y/x > y_i/x_i$$

and for those lying in the half plane

$$y/x < y_i/x_i.$$

If the disjunctive model applies, then for any individual such that $y/x > y_i/x_i$, he passes the item if $y \geq y_i$ and fails it if $y < y_i$. For any individual such that $y/x < y_i/x_i$, he passes the item if $x \geq x_i$ and fails it if $x < x_i$. In effect, we first state for any individual which of his two relevant abilities is "best" for the given item, and, secondly, whether he has enough of it.

One might say that the individual is assumed to utilize that relevant ability which is "best" for the given item.

The description just given immediately suggests an alternative though closely related model, which could also be regarded as disjunctive. Suppose that, for any individual such that $y > x$, he passes the item if $y \geq y_i$ and fails it if $y < y_i$. Conversely, for any individual such that $y < x$, he passes the item if $x \geq x_i$ and fails it if $x < x_i$. In effect, this would assume that throughout a series of items, the individual is set to utilize his best overall relevant ability, even though it may not be the best to use for certain specific items.

It is a matter of convention whether we regard the model just described as a second type of disjunctive model. In this, as in the case considered by Coombs, an individual's performance is a function of a "best relevant ability" only, and is independent of all other abilities. This seems to capture the essential notion. The difference is that in the model of the first kind, the ability utilized depends on both the item and the individual, while in the second kind it depends only on the individual. For these and other reasons that will appear, the Coombs' model will be referred to as the Type I disjunctive model, while the one just introduced will be called the Type II disjunctive model. The similarities and differences between the two models are illustrated by the diagrams in Fig. 3.4a.1, with five items labelled *a*, *b*, *c*, *d*, *e*.

Methods of analysis for the Type I model have been described by Coombs and others (cf. Torgerson, 1958). In the case of the Type II model, it is intuitively obvious that individuals in either half plane ($y > x$ or $y < x$) yield a Guttman scale if all x_i and all y_i are distinct. For n items there are $(n + 1)$ response patterns from each half plane. Two of these (failing all items and passing all items) are indistinguishable, while others may or may

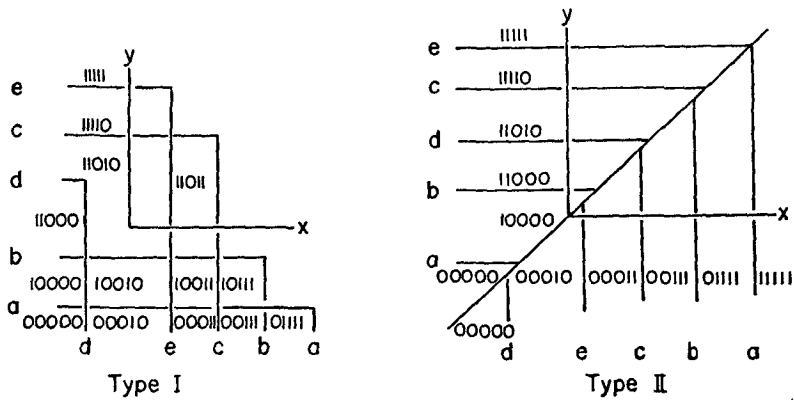


Figure 3.4a.1

not be, according to the relative orderings of the x_i and the y_i. Hence, in general, the Type II model yields m distinct response patterns where

$$n + 1 \leq m \leq 2n.$$

Thus, in principle, by an extension of scalogram techniques, one could divide the subjects into two groups giving distinct Guttman scales on the one set of items. It may also be noted that for given orderings of the x_i and y_i, the permissible response patterns for Type I include the permissible response patterns for Type II.

The two models as described can alternatively be specified in terms of regression functions (which are trace functions or conditional probability functions for the case of dichotomous observations) as follows:

As usual, we write*

$$\mathcal{E}(z_i \mid x, y) \underline{\equiv} \varphi_i(x, y)$$

for the regression function of the jth manifest variate on the latent variates. Then for either of the models under consideration, we have

$$\varphi_i(x, y) = u_i(x) \tag{3.4a.1}$$

in one of the defined half planes, and

$$\varphi_i(x, y) = v_i(y) \tag{3.4a.2}$$

in the other, with

$$\begin{aligned} u_i(x) &= 0, \quad x < x_i \\ &= 1, \quad x \geq x_i \end{aligned} \tag{3.4a.3}$$

and

$$\begin{aligned} v_i(y) &= 0, \quad y < y_i \\ &= 1, \quad y \geq y_i. \end{aligned} \tag{3.4a.4}$$

Within either half plane, the regression surface is a step function, taking values zero or unity. Considered in these terms, it could be said that Coombs' weak assumptions as to the metric of the factor space are purchased at the price of a quite strong restriction on the shape of the regression surface. This suggests, as a first basis for generalization, the substitution of other functions $u_i(x)$ and $v_i(y)$ for the regression surfaces in the respective half planes. We could consider, for example, polynomials, or step functions taking values other than zero or unity, as in the latent distance model.

The respective conditions defining the half planes in the two models can be thought of as conditions under which the individual utilizes the one

* Throughout this appendix we use z_j for the observed variate, *not necessarily* in standard form, to avoid confusion.

or the other ability. These conditions, like the regression functions, need not involve a discontinuity. As a second basis for generalization, we introduce the probabilities

$$n_i(x, y) \equiv \mathcal{P}\{\varphi_i(x, y) = u_i(x) \mid x, y\} \tag{3.4a.5}$$

and

$$\bar{n}_i(x, y) \equiv \mathcal{P}\{\varphi_i(x, y) = v_i(y) \mid x, y\}, \tag{3.4a.6}$$

where

$$n_i + \bar{n}_i = 1. \tag{3.4a.7}$$

The functions n_i and $\bar{n}_i$ will be referred to as *utilization* functions. The function n_i describes the conditional probability, given (x, y), that "ability x is utilized" and hence that $\varphi_i(x, y) = u_i(x)$. The function $\bar{n}_i$ is its converse and probability complement. To identify the Type II model it would be sufficient to assume that $n_i(x, y)$ is independent of j.

In the following the major results for the two models are presented first, then proofs of some of these are presented, in effect, as an appendix to this appendix.

General Regression Functions

On this basis only, without any alteration in the utilization functions, nothing worthwhile can be said at present about the Type I model. In the case of the Type II model, we might assume linear regressions in each half plane, viz.

$$\begin{aligned} u_i(x) &= a_{i0} + a_{i1}x \\ v_i(y) &= b_{i0} + b_{i1}y. \end{aligned} \tag{3.4a.8}$$

In this case it is easily shown that the covariance of two manifest variates z_i and z_k is given by

$$\operatorname{cov}(z_i, z_k) = p_x p_y (a_{i0} - b_{i0})(a_{k0} - b_{k0}) + p_x a_{i1} a_{k1} + p_y b_{i1} b_{k1}, \tag{3.4a.9}$$

where

$$\begin{aligned} p_x &= \mathcal{P}\{x > y\} \\ p_y &= \mathcal{P}\{y > x\}. \end{aligned} \tag{3.4a.10}$$

Thus, with suitable adjustments to the elements of the leading diagonal, this model yields a covariance matrix of rank three. It is of interest to consider some of the properties of the three obvious degenerate cases. If either p_x or p_y is zero, i.e., all subjects are on one or the other side of the line $y = x$, the model reduces as would be expected to a single-factor linear model. In case $a_{i0} = b_{i0}$ for all j, the covariance matrix can be accounted for by a

sum of products of the regression slopes within each half plane, weighted by the proportion of subjects therein. In case the slopes of the regression functions are zero, for all j, in both half planes, (3.4a.9) reduces to

$$\text{(3.4a.11)} \qquad \operatorname{cov}(z_j, z_k) = p_x p_y (a_{j0} - b_{j0})(a_{k0} - b_{k0})$$

and the model corresponds as we would expect, to the latent dichotomy model of sect. 4.3.

More generally, it can be shown that if the regression functions in each half plane can be represented in terms of orthogonal polynomials of the form

$$\text{(3.4a.12)} \qquad u_j(x) = \sum_{p=0}^{r} a_{jp} h_p(x)$$

$$v_j(y) = \sum_{s=0}^{t} b_{js} \ell_s(y),$$

then

$$\text{(3.4a.13)} \qquad \operatorname{cov}(z_j, z_k) = p_x p_y (a_{j0} - b_{j0})(a_{k0} - b_{k0}) + p_x \sum_{p=1}^{r} a_{jp} a_{kp} + p_y \sum_{s=1}^{t} b_{js} b_{ks}.$$

Hence it seems that this model, with two factors, will yield a covariance matrix of rank $(r + t + 1) \leq n$.

However, while the model as developed is not without conceptual interest, it seems unrealistic to assume a discontinuity at $y = x$ in the form of the regression functions. This is to suppose that the person correctly chooses to employ his better ability, no matter how slight its superiority. It is doubtful, then, whether it is worthwhile attempting to produce methods for fitting it to observed data. These results are offered only for their conceptual interest and their relations to the further generalizations below.

General Utilization Functions

Introducing the functions $n_j(x, y)$ and $\bar{n}_j(x, y)$ as in (3.4a.5) and (3.4a.6), we have immediately

$$\text{(3.4a.14)} \qquad \varphi_j(x, y) = u_j(x) n_j(x, y) + v_j(y) \bar{n}_j(x, y).$$

In general, the expression on the right in (3.4a.14) could be represented by a set of bivariate orthonormal functions in (x, y) i.e., by

$$\text{(3.4a.15)} \qquad \varphi_j(x, y) = \sum_{\ell=0}^{p} \sum_{m=0}^{q} a_{\ell m}^{(j)} h_{\ell m}(x, y),$$

where the $h_{\ell m}$ form an orthonormal set. We then have, in the usual way,

$$\text{(3.4a.16)} \qquad \operatorname{cov}(z_j, z_k) = \sum_{\ell=1}^{p} \sum_{m=1}^{q} a_{\ell m}^{(j)} a_{\ell m}^{(k)}.$$

In principle, data conforming to this model could be treated by nonlinear factor analysis, as in sect. 3.4, to yield regression functions in the form of (3.4a.15). While it is conceivable that the nature of these regression functions may suggest an obvious factorization of the type of (3.4a.14), such a factorization may not be unique.

For the corresponding properties of the Type II model we would substitute $n_i(x, y) = n(x, y)$ in the above relations. In the general case, this seems to make no useful difference.

In a following section we develop the particular case for both models in which the linear regression functions of (3.4a.8) are assumed. It is also assumed that the probability of utilization is a linear function of the perpendicular distance of (x, y) from the line

$$y/x = y_i/x_i.$$

With these assumptions, the Type I regression surface takes the form

$$\varphi_i(x, y) = \beta_{00}^{(i)} + \beta_{10}^{(i)}x + \beta_{01}^{(i)}y + \beta_{11}^{(i)}xy + \beta_{20}^{(i)}h_2(x) + \beta_{02}^{(i)}h_2(y), \tag{3.4a.17}$$

where in general the coefficients $\beta_{lm}^{(i)}$ are linearly independent, hence yielding a reduced covariance matrix of rank five. This function may be recognized as the quadric surface (2.5.5) discussed in sect. 2.5 and analyzed in sect. 3.4. Thus, if we assume linear functions where Coombs assumes step functions in the disjunctive model, we obtain a regression surface in the form of a general quadric. In principle, such a surface can be detected and fitted by nonlinear factor analysis.

With the same assumptions, the Type II regression surface is a degenerate case in which, in (3.4a.17),

$$\begin{aligned} \beta_{11}^{(i)} &= \gamma_1(\beta_{20}^{(i)} + \beta_{02}^{(i)}) \\ \beta_{10}^{(i)} + \beta_{01}^{(i)} &= \gamma_2(\beta_{20}^{(i)} + \beta_{02}^{(i)}), \end{aligned} \tag{3.4a.18}$$

where γ_1, γ_2 are constants. Hence this case in general yields a reduced covariance matrix of rank three. With the linear constraints (3.4a.18), it is instructive to transform (3.4a.17) into

$$\varphi_i(\xi, \eta) = \beta_{00}^{(i)} + \beta_{01}^{(i)}\eta + \beta_{11}^{(i)}\xi\eta + \beta_{10}^{(i)}\{\xi - \gamma h_2(\eta)\}, \tag{3.4a.19}$$

where γ is a constant. Given data conforming to this model, we might apply the methods of nonlinear factor analysis described earlier. Because of the nature of the three orthonormal components, a mechanical application of these methods could be expected to fail. It is possible, however, that inspection of the trivariate distribution of the sample component variates might lead us to suspect that this case applied.

There is a further sense in which the above considerations provide generalizations on the Coombs models. Throughout these developments we

have been concerned with regression functions $\varphi_i(x, y)$ which need not be regarded as the trace functions of dichotomous items. That is, the manifest variates as treated can be multivalued variables. If such is the case, the model correponds to the notion that an individual performs, on a given multivalued test, as well as his best relevant ability allows. This is in contrast to the more restricted notion that he passes a given item if his best relevant ability is sufficient.

Proof of (3.4a.9) *and* (3.4a.13)

Assuming that the latent variates (x, y) are statistically independent, we write for their joint density function

$$dF(x, y) = f(x)g(y)\,dx\,dy, \tag{3.4a.20}$$

taking nonzero values in the range $x_0 \leq x \leq x_1$, and $y_0 \leq y \leq y_1$.

We consider a set of n observed variates z_i (not necessarily in standard form). From the general principle of local independence (2.3.2), by (2.3.5), we have an expression for the expected value of any product of the observed variates

$$\mathcal{E}(z_i z_k \cdots z_n) = \int_{x=x_0}^{x=x_1} \int_{y=y_0}^{y=y_1} \varphi_i(x, y)\varphi_k(x, y) \cdots \varphi_n(x, y) f(x)g(y)\,dx\,dy, \tag{3.4a.21}$$

where for the case of the Type II model,

$$\begin{aligned} \varphi_i(x, y) &= u_i(x), \qquad y < x \\ &= v_i(y), \qquad y > x. \end{aligned} \tag{3.4a.22}$$

Hence (3.4a.21) becomes

$$\begin{aligned} \mathcal{E}(z_i z_k \cdots z_n) = &\int_{x=x_0}^{x=x_1} \int_{y=y_0}^{y=x} u_i(x)u_k(x) \cdots u_n(x) f(x)g(y)\,dx\,dy \\ &+ \int_{x=x_0}^{x=y} \int_{y=y_0}^{y=y_1} v_i(y)v_k(y) \cdots v_n(y) f(x)g(y)\,dx\,dy. \end{aligned} \tag{3.4a.23}$$

Writing

$$\begin{aligned} G(y) &= \int_{y_0}^{y} g(y)\,dy \\ F(x) &= \int_{x_0}^{x} f(x)\,dx \end{aligned} \tag{3.4a.24}$$

and noting that $F(x_0) = G(y_0) = 0$, (3.4a.23) becomes

$$\begin{aligned} \mathcal{E}(z_i z_k \cdots z_n) = &\int_{x=x_0}^{x=x_1} u_i(x)u_k(x) \cdots u_n(x) f_G(x)\,dx \\ &+ \int_{y=y_0}^{y=y_1} v_i(y)v_k(y) \cdots v_n(y) g_F(y)\,dy, \end{aligned} \tag{3.4a.25}$$

where

$$f_G(x) \equiv f(x)G(x) \tag{3.4a.26}$$
$$g_F(y) \equiv g(y)F(y).$$

We note that

$$\int_{x_0}^{x_1} f_G(x)\,dx + \int_{y_0}^{y_1} g_F(y)\,dy = \int_{x=x_0}^{x_1}\int_{y=y_0}^{y_1} f(x)g(y)\,dx\,dy = 1 \tag{3.4a.27}$$

so we write

$$\int_{x_0}^{x_1} f_G(x)\,dx = p_x, \tag{3.4a.28}$$
$$\int_{y_0}^{y_1} g_F(y)\,dy = p_y,$$

where

$$p_x + p_y = 1. \tag{3.4a.29}$$

The parameter p_x represents the proportion of the population in which $x > y$ and conversely for p_y.

The regression functions $u_i(x)$ and $v_i(y)$ can in general be represented in terms of a series of orthogonal functions whose norms are, respectively, p_x and p_y, i.e., we write

$$u_i(x) = \sum_{p=0}^{r} a_{ip}h_p(x) \tag{3.4a.30}$$
$$v_i(y) = \sum_{s=0}^{t} b_{is}\ell_s(y)$$

where

$$\int_{x_0}^{x_1} h_p(x)h_q(x)f_G(x)\,dx = p_x, \qquad p = q \tag{3.4a.31}$$
$$= 0, \qquad p \neq q$$
$$\int_{y_0}^{y_1} \ell_s(y)\ell_t(y)g_F(y)\,dy = p_y, \qquad s = t$$
$$= 0, \qquad s \neq t.$$

Then from (3.4a.25), (3.4a.30) and (3.4a.31), we have

$$\mathcal{E}(z_i) = p_x a_{i0} + p_y b_{i0} \tag{3.4a.32}$$

and

$$\mathcal{E}(z_i z_k) = p_x \sum_{p=0}^{r} a_{ip}a_{kp} + p_y \sum_{s=0}^{t} b_{is}b_{ks}. \tag{3.4a.33}$$

This yields in turn the required result (3.4a.13),

$$\text{cov}\,(z_j, z_k) = p_x p_y (a_{j0} - b_{j0})(a_{k0} - b_{k0}) + p_x \sum_{p=1}^{r} a_{jp} a_{kp} + p_y \sum_{s=1}^{t} b_{js} b_{ks},$$

by using (3.4a.32) and rearranging the zero-order terms with the assistance of (3.4a.29). If, in particular, the trace functions $u_i(x)$, $v_i(y)$ are assumed linear in x and y respectively, then (3.4a.13) reduces immediately to the other required result (3.4a.9) viz.

$$\text{cov}\,(z_j, z_k) = p_x p_y (a_{j0} - b_{j0})(a_{k0} - b_{k0}) + p_x a_{j1} a_{k1} + p_y b_{j1} b_{k1}.$$

Proof of (3.4a.17) *and* (3.4a.19)

Let us assume that

(3.4a.34)
$$u_i(x) = a_{i0} + a_{i1} x$$
$$v_i(y) = b_{i0} + b_{i1} y$$

and

(3.4a.35)
$$n_i = \tfrac{1}{2} + \alpha_i (y - \rho_i x)$$
$$\bar{n}_i = \tfrac{1}{2} - \alpha_i (y - \rho_i x),$$

where

(3.4a.36)
$$\alpha_i = \alpha \sqrt{\frac{\rho_i^2 + 1}{2\rho_i^2}},$$

α is an arbitrary parameter, and ρ_i corresponds to the ratio y_i/x_i in the discussion of Coombs' model above. The n_i and $\bar{n}_i$ then vary, in opposite senses, in proportion to the perpendicular distance of the point (x, y) from the line $y/x = y_i/x_i$, at a rate dependent on the constant α.

Substituting these expressions in (3.4a.14) yields

(3.4a.37)
$$\varphi_i(x, y) = \alpha_{00}^{(i)} + \alpha_{10}^{(i)} x + \alpha_{01}^{(i)} y + \alpha_{11}^{(i)} xy + \alpha_{20}^{(i)} x^2 + \alpha_{02}^{(i)} y^2,$$

where

(3.4a.38)
$$\alpha_{00}^{(i)} = \tfrac{1}{2}\{a_{i0} + b_{i0}\}$$
$$\alpha_{10}^{(i)} = \tfrac{1}{2}\{a_{i1} - \alpha_i \rho_i (a_{i0} - b_{i0})\}$$
$$\alpha_{01}^{(i)} = \tfrac{1}{2}\{b_{i1} + \alpha_i (a_{i0} - b_{i0})\}$$
$$\alpha_{11}^{(i)} = \tfrac{1}{2}\alpha_i \{a_{i1} + \rho_i b_{i1}\}$$
$$\alpha_{20}^{(i)} = -\tfrac{1}{2}\alpha_i \rho_i a_{i1}$$
$$\alpha_{02}^{(i)} = -\tfrac{1}{2}\alpha_i b_{i1},$$

and in general the coefficients in (3.4a.37) are linearly independent, given that the coefficients a_{i0}, b_{i0}, a_{i1}, b_{i1} are linearly independent.

For any prescribed density function $f(x)g(y)$ of the statistically independent latent variates x and y, (3.4a.37) can be transformed into the required form of (3.4a.17)

$$\varphi_i(x, y) = \beta_{00}^{(i)} + \beta_{10}^{(i)}x + \beta_{01}^{(i)}y + \beta_{11}^{(i)}xy + \beta_{20}^{(i)}h_2(x) + \beta_{02}^{(i)}h_2(y).$$

The Type II model is a degenerate case of the Type I model in which

$$\rho_i = 1, \qquad \alpha_i = \alpha,$$
$$n_i = n = \tfrac{1}{2} + \alpha(y - x),$$
$$\bar{n}_i = \bar{n} = \tfrac{1}{2} - \alpha(y - x).$$

Substituting these in (3.4a.38), it is evident by inspection in this case that α_{11} and $\alpha_{10} + \alpha_{01}$ are linear functions of α_{20} and α_{02}.

Introducing the transformation

$$\xi = \frac{1}{\sqrt{2}}(x + y)$$

$$\eta = \frac{1}{\sqrt{2}}(x - y),$$

with the assumptions of the Type II model, (3.4a.37) readily yields

(3.4a.39) $$\varphi_i(\xi, \eta) = \alpha_{00}^{(i)} + \alpha_{01}^{(i)}\eta + \alpha_{10}^{(i)}\{\xi - \gamma\eta^2\} + \alpha_{11}^{(i)}\xi\eta,$$

where

$$\alpha_{00}^{(i)} = \tfrac{1}{2}(a_{i0} + b_{i0})$$

$$\alpha_{01}^{(i)} = \frac{1}{2\sqrt{2}}\{a_{i1} - b_{i1} - 2\alpha(a_{i0} - b_{i0})\}$$

$$\alpha_{10}^{(i)} = \frac{1}{2\sqrt{2}}(a_{i1} + b_{i1})$$

$$\alpha_{11}^{(i)} = -\tfrac{1}{2}\alpha(a_{i1} - b_{i1})$$

and

$$\gamma = (2\sqrt{2})\,\alpha.$$

Given prescribed density functions for $f(x)$, $g(y)$, (3.4a.39) is readily transformed into the required result (3.4a.19).

APPENDIX 4.3A

Proofs of Results in Section 4.3

(Latent Class Model)

(1) Proof of the relations between the parameters of the latent linear model and those of the latent dichotomy:

In the latent linear model

$$\mathcal{E}(y_i \mid x) \equiv \varphi_i(x) = a_i + b_i x, \tag{4.3a.1}$$

let x have a discrete distribution, taking the values x_1, x_2 with probabilities g_1, g_2 ($x_1 < x_2$, $g_1 + g_2 = 1$). If x is in standard form,

$$\mathcal{E}(x) = (g_1 x_1 + g_2 x_2) = 0 \tag{4.3a.2}$$

and

$$\text{var}\,(x) = (g_1 x_1^2 + g_2 x_2^2) = 1, \tag{4.3a.3}$$

whence immediately

$$x_1 = -\sqrt{\frac{g_2}{g_1}} \tag{4.3a.4}$$

$$x_2 = \sqrt{\frac{g_2}{g_2}}.$$

Equation (4.3a.1) then yields

$$\varphi_{i1} \equiv \varphi_i(x_1) = a_i - b_i\sqrt{g_2/g_1}, \tag{4.3a.5}$$

and

$$\varphi_{i2} \equiv \varphi_i(x_2) = a_i + b_i\sqrt{g_1/g_2}. \tag{4.3a.6}$$

Solving (4.3a.5) and (4.3a.6) for a_i and b_i yields

$$a_i = \varphi_{i1} g_1 + \varphi_{i2} g_2 \tag{4.3a.7}$$

and

$$b_i = \sqrt{g_1 g_2}(\varphi_{i2} - \varphi_{i1}). \tag{4.3a.8}$$

(2) Proof of the result for the variance of the disturbance component.

From (4.3a.1) we have

$$\text{cov}\,(y_i, y_k) = b_i b_k.$$

By (4.3.6), this yields in turn

$$\text{cov}\,(y_i, y_k) = g_1 g_2 \delta_i \delta_k.$$

In this case, then, the factor matrix F takes the form

$$F = [\sqrt{g_1 g_2}\ \delta_1 \cdots \sqrt{g_1 g_2}\ \delta_n]$$

and the diagonal matrix C of nonzero latent roots reduces to the scalar

$$C = g_1 g_2 \sum_{i=1}^{n} \delta_i^2.$$

Further, since the variance of any observed variate y_i is given by

$$\text{var}\,(y_i) = a_i(1 - a_i),$$

the diagonal matrix U of coefficients of unique variation is made up of elements

$$u_i = \sqrt{a_i(1 - a_i) - g_1 g_2 \delta_i^2}.$$

Substituting accordingly for F, U, and C in the expression for the covariance matrix S of the disturbance components given by (3.1.24) and (3.1.25) yields the scalar

$$\text{var}\,(d) \equiv \underline{S} = \frac{\sum_{i=1}^{n} \delta_i^2 \{a_i(1 - a_i) - g_1 g_2 \delta_i^2\}}{g_1 g_2 \left(\sum_{i=1}^{n} \delta_i^2\right)^2}.$$

If we assume that $\varphi_i = \varphi$ for all j, and $a_j = g_1$ for all j, this expression reduces to

$$\text{var}\,(d) = \frac{1 - \delta^2}{n\delta^2}.$$

APPENDIX 4.3B

Covariance Matrix

(.162)	.139	.128	.122	.116	.120	.114	.157	.152	.124	−.015	−.022	−.008	−.039	.019	.019	−.015	−.011	.009	−.027
.130	(.120)	.110	.105	.100	.103	.098	.135	.131	.107	−.012	−.019	−.007	−.034	.016	.017	−.013	−.010	.008	−.023
.145	.085	(.100)	.096	.091	.094	.090	.124	.120	.098	−.011	−.017	−.006	−.031	.015	.015	−.012	−.009	.007	−.021
.115	.115	.101	(.092)	.088	.091	.086	.119	.115	.094	−.011	−.016	−.006	−.030	.014	.014	−.011	−.009	.007	−.020
.120	.110	.086	.104	(.083)	.086	.082	.112	.109	.089	−.010	−.016	−.006	−.028	.014	.014	−.011	−.008	.007	−.019
.115	.105	.090	.110	.095	(.089)	.084	.116	.113	.092	−.011	−.016	−.006	−.029	.014	.014	−.011	−.008	.007	−.020
.120	.110	.105	.075	.070	.075	(.080)	.110	.107	.087	−.010	−.015	−.006	−.028	.013	.013	−.010	−.008	.007	−.019
.184	.124	.121	.128	.113	.110	.106	(.152)	.147	.120	−.014	−.021	−.008	−.038	.018	.018	−.014	−.011	.009	−.026
.155	.125	.130	.099	.104	.120	.105	.149	(.143)	.116	−.014	−.020	−.008	−.037	.018	.018	−.014	−.011	.009	−.025
.115	.125	.080	.079	.064	.100	.115	.109	.140	(.095)	−.011	−.017	−.006	−.030	.014	.014	−.011	−.009	.007	−.021
.011	−.009	−.026	−.002	−.038	−.025	.009	.019	−.014	.016	(.001)	.002	.001	.004	−.002	− 002	.001	.001	−.001	.002
−.019	−.019	−.026	.008	−.018	.005	−.031	−.021	−.004	−.024	−.015	(.003)	.001	.005	−.003	−.003	.002	.002	−.001	.004
−.015	−.005	.000	.009	.004	.000	.015	−.001	−.030	.010	−.054	−.024	(.000)	.002	−.001	−.001	.001	.001	−.000	.001
−.030	−.030	−.004	−.046	−.021	−.015	−.040	−.027	−.046	−.016	.032	.032	.004	(.010)	−.005	−.005	.004	.003	−.002	.007
.031	.041	.033	.019	.003	−.005	.029	.002	−.003	.027	−.028	−.018	−.013	−.007	(.002)	.002	−.002	−.001	.001	−.003
.025	.015	.020	.010	.035	.030	.005	.030	.010	.010	.015	−.025	.000	.005	.015	(.002)	−.002	−.001	.001	−.003
.010	−.040	.046	−.006	−.021	−.015	.010	−.007	−.016	−.027	.012	−.018	.004	.029	.003	.015	(.001)	.001	−.001	.002
−.005	−.015	.031	−.022	−.046	.010	−.015	.008	−.011	.009	.028	−.012	.009	.004	.039	.040	−.006	(.001)	−.001	.002
.005	.005	.030	.009	−.016	−.020	−.005	.019	.040	.030	.006	−.004	−.030	.024	−.003	− 040	.004	−.021	(.001)	−.002
.000	−.030	−.016	−.044	−.039	−.015	−.020	−.003	−.014	−.014	−.002	−.092	.026	.009	.037	.005	.021	.036	.016	(.004)

Lower triangle, observed covariances
Upper triangle, reproduced covariances
Leading diagonal, communalities

APPENDIX 4.4

Proof of the Relations in Section 4.4

(*Equations* 4.4.1 *and* 4.4.5)

(1) *Derivation of Equation* 4.4.1

In this case, I evaluated separately the first four coefficients, c_{j0}, c_{j1}, c_{j2} and c_{j3} given explicitly in (4.4.1) and (4.4.2). This led to a conjecture as to the general result for c_{jp}, which was as it happens correct. An elegant proof of the general formula for all $p \geq 1$ was kindly supplied by Dr. E. J. Burr of the Department of Mathematics. The special case of c_{j0} will be given first, and this is followed by the proof supplied to me for the general case. In the following it is convenient to omit the subscript j.

We wish to evaluate the integral

$$c_p = \int_{-\infty}^{\infty} \varphi(x) h_p(x) g(x)\, dx, \tag{4.4a.1}$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} \exp\left(-\frac{z^2}{2}\right) dz$$

$$h_p(x) = \frac{1}{\sqrt{p!}} (-)^p \exp\left(\frac{x^2}{2}\right) \frac{d^p}{dx^p} \exp\left(-\frac{x^2}{2}\right)$$

$$g(x) = n(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Case of c_0

For $p = 0$, (4.4a.1) yields

$$c_0 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} \exp\left(-\frac{z^2}{2}\right) dz \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. \tag{4.4a.2}$$

We make the variate transformation

$$[u, v] = \left[\frac{\sigma x + z}{\sqrt{1+\sigma^2}}, \frac{-x + \sigma z}{\sqrt{1+\sigma^2}}\right] \tag{4.4a.3}$$

with the inverse

$$[x, z] = \left[\frac{\sigma u - v}{\sqrt{1+\sigma^2}}, \frac{u + \sigma v}{\sqrt{1+\sigma^2}}\right] \tag{4.4a.4}$$

which, being a rotation, has a Jacobian of unity.

This transforms the line $z = (x - \mu)/\sigma$ into the line $v = -\mu/\sqrt{1 + \sigma^2}$, and we have

$$(4.4a.5) \qquad c_{j0} = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{-\mu/\sqrt{1+\sigma^2}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(u+\sigma v)^2}{1+\sigma^2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(\sigma u - v)^2}{1+\sigma^2}\right\} du\, dv$$

$$= \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{-\mu/\sqrt{1+\sigma^2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv$$

$$= \frac{1}{\sqrt{2\pi}} \int_{v=-\infty}^{-\mu/\sqrt{1+\sigma^2}} \exp\left(-\frac{v^2}{2}\right) dv$$

$$= N\{-\mu/\sqrt{1+\sigma^2}\}.$$

Using the same variate transformation, it was possible to evaluate c_{j1}, c_{j2} and c_{j3}, and hence to conjecture the general formula. However, proceeding in this way yields successively more tedious problems in elementary calculus.

General Result for $p \geq 1$

We define the set of polynomials $H_p(x)$ by

$$(4.4a.6) \qquad H_p(x) \exp\left(-\frac{x^2}{2}\right) = \left(-\frac{d}{dx}\right)^p \exp\left(-\frac{x^2}{2}\right), \qquad p = 0, 1, 2 \cdots$$

whence

$$(4.4a.7) \qquad H_{p+1}(x) \exp\left(-\frac{x^2}{2}\right) = -\frac{d}{dx}\left[H_p(x) \exp\left(-\frac{x^2}{2}\right)\right].$$

The Fourier transform of $g(x)$ is defined as

$$(4.4a.8) \qquad G(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) \exp(i\theta x)\, dx$$

and has the property that

$$(4.4a.9) \qquad g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(\theta) \exp(-i\theta x)\, d\theta.$$

It is well known that

$$(4.4a.10) \qquad \exp(-\tfrac{1}{2}\lambda^2\theta^2), \qquad \frac{1}{\lambda} \exp\left(-\frac{1}{2}\frac{x^2}{\lambda^2}\right), \qquad \lambda > 0,$$

are Fourier transforms of one another. Also

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(i\theta)^p \exp(-\tfrac{1}{2}\lambda^2\theta^2)\exp(-i\theta x)\,d\theta$$

$$=\frac{1}{\sqrt{2\pi}}\left(-\frac{d}{dx}\right)^p\int_{-\infty}^{\infty}\exp(-\tfrac{1}{2}\lambda^2\theta^2)\exp(-i\theta x)\,d\theta.$$

By (4.4a.9) and (4.4a.10),

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(i\theta)^p \exp(-\tfrac{1}{2}\lambda^2\theta^2)\exp(-i\theta x)\,d\theta = \left(-\frac{d}{dx}\right)^p\left[\frac{1}{\lambda}\exp\left(-\frac{1}{2}\frac{x^2}{\lambda^2}\right)\right].$$

Hence, by (4.4a.6),

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(i\theta)^p \exp(-\tfrac{1}{2}\lambda^2\theta^2)\exp(-i\theta x)\,d\theta = \left(\frac{1}{\lambda}\right)^{p+1}H_p\left(\frac{x}{\lambda}\right)\exp\left(-\frac{1}{2}\frac{x^2}{\lambda^2}\right).$$

That is, the Fourier transform of

(4.4a.11) $$\frac{1}{\lambda^{p+1}}H_p\left(\frac{x}{\lambda}\right)\exp\left(-\frac{1}{2}\frac{x^2}{\lambda^2}\right) \quad \text{is} \quad (i\theta)^p\exp(-\tfrac{1}{2}\lambda^2\theta^2).$$

Evaluation of c_p $(p \geq 1)$

We write

$$\sqrt{(p+1)!}\,c_{p+1} \equiv \underline{g(\mu)}.$$

Then by definition of c_p,

$$g(\mu) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\varphi(x)H_{p+1}(x)\exp\left(-\frac{x^2}{2}\right)dx,$$

and by (4.4a.7)

(4.4a.12) $$g(\mu) = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\varphi(x)\,d\left[H_p(x)\exp\left(-\frac{x^2}{2}\right)\right]$$

$$= +\frac{1}{2\pi\sigma}\int_{-\infty}^{\infty}H_p(x)\exp\left(-\frac{x^2}{2}\right)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx$$

on integration by parts, since

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

We now find the Fourier transform of $g(\mu)$.

$$\underline{G(\theta)} \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}g(\mu)\exp(i\mu\theta)\,d\theta$$

$$= \frac{1}{(2\pi)^{3/2}\sigma} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ H_p(x) \exp\left(-\frac{x^2}{2}\right) \exp(ix\theta) \right\}$$

$$\cdot \left\{ \exp\left(-\frac{(\mu - x)^2}{2\sigma^2}\right) \exp i(\mu - x)\theta \right\} dx\, du$$

$$= \frac{1}{\sqrt{2\pi}} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_p(x) \exp\left(-\frac{x^2}{2}\right) \exp(ix\theta)\, dx \right]$$

$$\cdot \left[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{y^2}{\sigma^2}\right) \exp(iy\theta)\, dy \right],$$

where

$$y = \mu - x.$$

Hence by (4.4a.10) and (4.4a.11)

$$G(\theta) = \frac{1}{\sqrt{2\pi}} \left[(i\theta)^p \exp\left(-\frac{\theta^2}{2}\right) \right] [\exp(-\tfrac{1}{2}\sigma^2\theta^2)] \tag{4.4a.13}$$

$$= \frac{1}{\sqrt{2\pi}} (i\theta)^p \exp(-\tfrac{1}{2}\lambda^2\theta^2),$$

where

$$\lambda \equiv \sqrt{1 + \sigma^2}.$$

(Alternatively, the convolution theorem applied to (4.4a.12) yields (4.4a.13) immediately.)

Hence by taking inverse Fourier transforms, using (4.4a.11),

$$g(\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\lambda^{p+1}} H_p\left(\frac{\mu}{\lambda}\right) \exp\left(-\frac{1}{2}\frac{\mu^2}{\lambda^2}\right).$$

Therefore

$$c_p = \frac{1}{\sqrt{2\pi p!}} \lambda^{-p} H_{p-1}\left(\frac{\mu}{\lambda}\right) \exp\left(-\frac{1}{2}\frac{\mu^2}{\lambda^2}\right) \qquad p = 1, 2, \cdots,$$

that is,

$$c_p = \frac{1}{\sqrt{p}} (1 + \sigma^2)^{-p/2} h_{p-1}\{\mu/(1 + \sigma^2)^{1/2}\} n\{\mu/(1 + \sigma^2)^{1/2}\}.$$

(2) *Derivation of Equation* 4.4.5

In the conditional probability function

$$\varphi_i(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu_i)/\sigma_i} \exp\left(-\frac{z^2}{2}\right) dz,$$

$$\varphi_i(x) \to 1, \qquad x \to \infty$$

$$\to 0, \qquad x \to -\infty.$$

We obtain a corresponding function $\varphi_i^*(x)$ such that

$$\varphi_i^*(x) \to \lambda_{i1}, \qquad x \to \infty$$
$$\varphi_i^*(x) \to \lambda_{i0}, \qquad x \to -\infty$$

simply by defining

$$\varphi_i^*(x) \equiv \underline{\lambda_{i0}} + (\lambda_{i1} - \lambda_{i0})\varphi_i(x).$$

Writing for the moment

$$c_{ip}^* = \int_{-\infty}^{\infty} \varphi_i(x)h_p(x)n(x)\,dx$$

in the notation employed above, we have immediately

$$c_{ip}^* = \lambda_{i0} \int_{-\infty}^{\infty} h_p(x)n(x)\,dx + (\lambda_{i1} - \lambda_{i0})c_{ip}.$$

Given that

$$\int_{-\infty}^{\infty} h_p(x)n(x)\,dx = 1, \qquad p = 0$$
$$= 0, \qquad p = 1, 2, \cdots$$

we have

$$c_{i0}^* = \lambda_{i0} + (\lambda_{i1} - \lambda_{i0})c_{i0}$$
$$c_{ip}^* = (\lambda_{i1} - \lambda_{i0})c_{ip}, \qquad p = 1, 2, \cdots$$

so that, reverting to the notation of the main text,

(4.4.5) $$c_{i0} = \lambda_{i0} + (\lambda_{i1} - \lambda_{i0})N\{-\mu_i/(1 + \sigma_i^2)^{1/2}\}$$
$$c_{ip} = (\lambda_{i1} - \lambda_{i0})p^{-1/2}(1 + \sigma^2)^{-p/2}h_{p-1}\{\mu_i/(1 + \sigma^2)^{1/2}\}$$
$$\cdot n\{\mu_i/(1 + \sigma_i^2)^{1/2}\}, \qquad p = 1, 2, \cdots.$$

APPENDIX 4.5

Proof of the Relation in Section 4.5 (Equation 4.5.1)

Given the conditional probability function

$$\varphi_i(x) = \gamma_i, \qquad x \leq x_i \tag{4.5a.1}$$
$$= \beta_i, \qquad x > x_i,$$

we may rewrite this in the form

$$\varphi_i(x) = \beta_i + (\gamma_i - \beta_i) f_i(x), \tag{4.5a.2}$$

where

$$f_i(x) = 1, \qquad x \leq x_i \tag{4.5a.3}$$
$$= 0, \qquad x > x_i.$$

It is then sufficient to determine the coefficients α_{ip} in the Fourier expansion

$$f_i(x) \sim \sum_{p=0} \alpha_{ip} h_p(x) \tag{4.5a.4}$$

since the required Fourier coefficients, c_{ip}, for $\varphi_i(x)$ will be given by

$$c_{ip} = \int_{-\infty}^{\infty} \varphi_i(x) h_p(x) n(x)\, dx$$
$$= \beta_i \int_{-\infty}^{\infty} h_p(x) n(x)\, dx + (\gamma_i - \beta_i) \int_{-\infty}^{\infty} f_i(x) h_p(x) n(x)\, dx,$$

whence

$$c_{i0} = \beta_i + (\gamma_i - \beta_i)\alpha_{i0} \tag{4.5a.5}$$
$$c_{ip} = (\gamma_i - \beta_i)\alpha_{ip}, \qquad p = 1, 2, \cdots .$$

Given that

$$\alpha_{ip} = \int_{-\infty}^{\infty} f_i(x) h_p(x) n(x)\, dx, \tag{4.5a.6}$$

where $f_i(x)$ is defined by (4.5a.3),

$$h_p(x) = \frac{1}{\sqrt{p!}} (-)^p \exp(\tfrac{1}{2}x^2) \frac{d^p}{dx^p} \exp(-\tfrac{1}{2}x^2)$$

and

$$n(x) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}x^2),$$

we have

$$\alpha_{ip} = \int_{-\infty}^{x_i} \frac{1}{\sqrt{p!}} (-)^p \exp(\tfrac{1}{2}x^2) \frac{d^p}{dx^p} \exp(-\tfrac{1}{2}x^2) \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}x^2)\, dx \tag{4.5a.7}$$

$$= \frac{1}{\sqrt{2\pi p!}} (-)^p \int_{-\infty}^{x_i} \frac{d^p}{dx^p} \exp(-\tfrac{1}{2}x^2)\, dx.$$

This yields immediately

$$\alpha_{i0} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i} \exp(-\tfrac{1}{2}x^2)\, dx = N(x_i), \tag{4.5a.8}$$

and

$$\alpha_{ip} = \frac{1}{\sqrt{2\pi p!}} \left[(-)^p \frac{d^{p-1}}{dx^{p-1}} \exp(-\tfrac{1}{2}x^2) \right]_{-\infty}^{x_i} \tag{4.5a.9}$$

$$= -\frac{1}{\sqrt{p}} h_{p-1}(x_i) n(x_i), \qquad p = 1, 2, \cdots.$$

Finally, by (4.5a.5) we have

$$c_{i0} = \beta_i + (\gamma_i - \beta_i) N(x_i)$$

$$c_{ip} = \frac{1}{\sqrt{p}} (\beta_i - \gamma_i) h_{p-1}(x_i) n(x_i), \qquad p = 1, 2, \cdots.$$

APPENDIX 4.7A

Properties of the Model (4.7.4)

We have

$$y_i = f_i(x), \tag{4.7.4}$$

where

$$\begin{aligned} f_i(x) &= 0, \qquad x \leq x_{i0} \\ &= \varphi_i(x), \qquad x_{i0} < x < x_{i1} \\ &= 1, \qquad x \geq x_{i1}. \end{aligned}$$

We first show that this model yields the property (2.3.22)

$$r_{ik} = \alpha_i/\alpha_k, \qquad k > j$$

if and only if $\varphi_i(x)$ is a step-function.

Consider two observed variates j, k such that $x_{k0} > x_{j1}$. Writing as usual $g(x)$ for the density function of the latent variate x, we have by (4.7.4)

$$\mathcal{E}(y_j) = \int_{x_{j0}}^{x_{j1}} \varphi_j(x)g(x)\,dx + \int_{x_{j1}}^{\infty} g(x)\,dx. \tag{4.7a.1}$$

We introduce

$$\begin{aligned} J &\equiv \mathcal{E}(y_j) \\ \Phi_j &\equiv \int_{x_{j0}}^{x_{j1}} \varphi_j(x)g(x)\,dx \\ G_j &\equiv \int_{x_{j1}}^{\infty} g(x)\,dx, \end{aligned} \tag{4.7a.2}$$

with analogous notation for items k, ℓ, so that

$$J = \Phi_j + G_j.$$

Similarly, we have

$$\mathcal{E}(y_j^2) = \int_{x_{j0}}^{x_{j1}} \{\varphi_j(x)\}^2 g(x)\,dx + \int_{x_{j1}}^{\infty} g(x)\,dx. \tag{4.7a.3}$$

We introduce

(4.7a.4) $$J_2 \equiv \mathcal{E}(y_i^2)$$

$$\Phi_{i2} \equiv \int_{x_{i0}}^{x_{i1}} \{\varphi_i(x)\}^2 g(x)\, dx.$$

Then by (4.7a.2), (4.7a.3) and (4.7a.4) we have

(4.7a.5) $$J_2 = \Phi_{i2} + G_i.$$

We write

(4.7a.6) $$\Phi_i = \Phi_{i2} + \theta_i,$$

where $\theta_i \geq 0$. Then by (4.7a.5) and (4.7a.6)

(4.7a.7) $$J_2 = J - \theta_i,$$

and

(4.7a.8) $$\begin{aligned} \text{var}\,(y_i) &= J - \theta_i - J^2 \\ &= J(1 - J) - \theta_i. \end{aligned}$$

Further, by (4.7.4) and (4.7a.2) we have

(4.7a.9) $$\mathcal{E}(y_i y_k) = \mathcal{E}(y_k) = K, \qquad (y_{k0} > y_{i1}).$$

This yields in turn

(4.7a.10) $$\text{cov}\,(y_i, y_k) = K(1 - J),$$

so that we have finally an expression for the correlation coefficient from (4.7a.8) and (4.7a.10), viz.

(4.7a.11) $$r_{ik} = \frac{K(1 - J)}{\sqrt{\{K(1 - K) - \theta_k\}\{J(1 - J) - \theta_i\}}},$$

$$x_{k0} > x_{i1}.$$

Rewriting this last in the form

(4.7a.12) $$r_{ik} = \frac{1 - J}{\sqrt{J(1 - J) - \theta_i}} \Big/ \frac{\sqrt{K(1 - K) - \theta_k}}{K}$$

and noting that similarly

(4.7a.13) $$r_{k\ell} = \frac{1 - K}{\sqrt{K(1 - K) - \theta_k}} \Big/ \frac{\sqrt{L(1 - L) - \theta_\ell}}{L},$$

$$x_{\ell 0} > x_{k1},$$

we see that for the property $r_{jk} = a_j/a_k$ to hold, for all j, k, we must have

$$\frac{\sqrt{K(1-K)-\theta_k}}{K} = \frac{1-K}{\sqrt{K(1-K)-\theta_k}}. \tag{4.7a.14}$$

It follows immediately that we must have $\theta_k = 0$ for all k.

Now by (4.7a.2), (4.7a.4) and (4.7a.6), we have

$$\begin{aligned}\theta_k &= \int_{x_{k0}}^{x_{k1}} \{\varphi_k(x)\} g(x)\, dx - \int_{x_{k0}}^{x_{k1}} \{\varphi_k(x)\}^2 g(x)\, dx \\ &= \int_{x_{k0}}^{x_{k1}} \varphi_k(x)\{1-\varphi_k(x)\} g(x)\, dx,\end{aligned} \tag{4.7a.15}$$

where $0 \leq \varphi_k(x) \leq 1$. Hence, $\theta_k = 0$ if and only if $\varphi_k(x)$ is equal to zero or unity almost everywhere in the range $x_{k0} \leq x \leq x_{k1}$. If x is continuously measurable, $\varphi_k(x)$ is a step-function taking the values zero and unity.

We turn now to the assertion in sect. 4.7 that the model (4.7.4) yields a quasi-simplex. The formal basis for this assertion is the form of the ratio in (4.7a.12). The expression (4.7a.12) reduces to the simplex property (2.3.22) if and only if $\theta_j = 0$ for all j. Provided, then, that θ_j is "small" for all j, it should prove difficult to distinguish the situation represented by (4.7a.12) from the situation represented by (2.3.22), on the basis of a finite sample of data. It is in this sense that the property (4.7a.12) can be considered to define a *quasi-simplex.*

APPENDIX 4.7B

The Translated Quadratic Model

From the general expression for a "translated function" model, viz.

$$\varphi_i(x) = \varphi(x - x_i), \tag{4.7.5}$$

we obtain the case where φ is a quadratic function in x, whence we have

$$y_i = a(x - x_i)^2 + b(x - x_i) + c, \tag{4.7b.1}$$

where a, b and c are constants. Writing

$$k = 1/\sqrt{\mu_4 - \mu_3^2 - 1}$$

where μ_3, μ_4 are the third and fourth moments of the distribution of x, we may rewrite (4.7b.1) in the form

$$y_i = \frac{a}{k} h_2(x) + \{b - a(2x_i - \mu_3)\}x + a(x_i^2 - 1) - bx_i + c. \tag{4.7b.2}$$

From (4.7b.2) we obtain

$$\mathcal{E}(y_i) = a(x_i^2 - 1) - bx_i + c, \tag{4.7b.3}$$

$$\mathcal{E}(y_i^2) = \frac{a^2}{k^2} + \{b - a(2x_i - \mu_3)\}^2 + \{a(x_i^2 - 1) - bx_i + c\}^2 \tag{4.7b.4}$$

and

$$\begin{aligned}\mathcal{E}(y_i y_k) = {} & \frac{a^2}{k} + \{b - a(2x_i - \mu_3)\}\{b - a(2x_k - \mu_3)\} \\ & + \{a(x_i^2 - 1) - bx_i + c\}\{a(x_k^2 - 1) - bx_k + c\}.\end{aligned} \tag{4.7b.5}$$

Writing

$$\begin{aligned}\alpha &= b + a\mu_3 \\ \beta &= -2a \\ \gamma &= a/k,\end{aligned} \tag{4.7b.6}$$

we obtain from (4.7b.3), (4.7b.4) and (4.7b.5),

$$r_{ik} = \frac{\gamma^2 + (\alpha + \beta x_i)(\alpha + \beta x_k)}{\sqrt{\{\gamma^2 + (\alpha + \beta x_i)^2\}\{\gamma^2 + (\alpha + \beta x_k)^2\}}}. \tag{4.7b.7}$$

For the numerical example given in sect. 4.7 we substitute in (4.7b.7) the values $\alpha = 0, \beta = 1, \gamma = 1$, corresponding, say, to $a = \sqrt{2}, b = 0$, in (4.7b.1), with a normal density function for x.

This yields

$$r_{ik} = \frac{1 + x_i x_k}{\sqrt{(1 + x_i^2)(1 + x_k^2)}}$$

whence in turn we obtain the matrix of Table 4.7.1 in which successive rows or columns correspond respectively to $x_i = -2, -1, 0, 1$ and 2.

APPENDIX 5.2

Covariance Matrix

(.029)	.051	.050	.014	.001	.016	.008	.011	.004	.019	.040	.031	.057	.065	.026	.014	.017	.015
.047	(.106)	.095	−.017	−.022	.003	−.007	.002	−.010	.031	.066	.049	.099	.117	.018	−.007	.007	.003
.054	.100	(.090)	.006	−.007	.014	.002	.007	.000	.032	.070	.053	.099	.114	.024	.006	.011	.015
.006	−.014	.016	(.122)	.069	.071	.056	.041	.046	.016	.043	.032	.037	.029	.066	.083	.050	.066
.003	−.016	.003	.070	(.041)	.037	.030	.021	.026	.005	.017	.013	.009	.002	.030	.044	.023	.035
.011	.007	.013	.086	.027	(.052)	.042	.037	.028	.013	.026	.021	.031	.033	.070	.064	.056	.044
.013	.003	.007	.057	.038	.050	(.035)	.031	.022	.008	.012	.010	.014	.015	.059	.053	.049	.034
.017	−.001	.011	.043	.015	.045	.037	(.030)	.017	.008	.009	.008	.015	.018	.061	.048	.051	.029
.006	−.005	.004	.038	.044	.024	.026	.019	(.018)	.005	.013	.010	.010	.007	.028	.033	.022	.025
.035	.031	.034	.017	.002	.010	.007	.009	.003	(.013)	.028	.021	.038	.043	.017	.012	.011	.013
.027	.069	.048	.038	.010	.025	.006	.012	.009	.031	(.072)	.053	.089	.094	.017	.019	.005	.028
.054	.046	.043	.029	.009	.018	.003	.016	.005	.023	.067	(.040)	.067	.071	.016	.016	.006	.022
.050	.081	.089	.030	.006	.035	.013	.008	.006	.039	.099	.064	(.118)	.130	.035	.023	.018	.031
.062	.116	.115	.032	.001	.030	.012	.016	.009	.029	.088	.052	.143	(.149)	.049	.026	.030	.031
.027	.005	.025	.068	.023	.076	.050	.065	.022	.022	.026	.019	.041	.040	(.131)	.093	.110	.052
.014	−.005	.004	.075	.056	.040	.055	.029	.040	.014	.019	.014	.021	.033	.086	(.082)	.077	.052
.012	.016	.000	.047	.026	.055	.050	.050	.022	.007	.004	.011	.018	.035	.096	.092	(.094)	.041
.025	−.003	.012	.059	.026	.042	.024	.031	.028	.013	.029	.018	.030	.035	.046	.066	.037	(.039)

Lower triangle, observed covariances
Upper triangle, reproduced covariances
Leading diagonal, communalities

APPENDIX 5.8

TABLE 5.8A.1
Score Matrix

Trial	Subjects																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	9	16	10	14	7	7	10	11	7	7	8	12	12	5	11	8	10	11
2	16	19	12	19	12	7	13	13	11	12	14	19	11	12	8	18	11	14
3	16	17	14	14	15	7	15	13	7	12	6	14	15	13	12	7	12	17
4	14	11	14	18	15	11	19	11	12	12	12	19	20	15	11	11	12	18
5	17	21	20	18	17	11	22	15	12	15	17	20	15	8	13	19	18	19
6	18	21	21	18	19	11	22	15	7	15	14	23	14	14	14	14	19	21
7	18	19	21	19	16	16	23	16	12	3	15	26	15	13	12	15	20	21
8	18	19	23	23	20	16	25	16	15	10	21	23	15	20	17	17	20	22
9	23	21	23	21	15	18	26	17	13	12	20	29	17	23	15	20	21	23
10	24	19	24	23	21	18	26	19	14	13	14	29	20	17	17	20	23	23
11	25	23	19	25	19	21	27	13	17	13	21	31	20	18	17	23	22	20
12	26	23	25	21	17	21	25	17	15	14	21	28	26	17	17	25	23	23
13	26	23	27	26	17	25	27	13	16	14	21	31	25	20	20	25	23	26
14	26	23	28	25	21	25	28	19	17	14	21	33	27	20	21	25	24	29
15	29	25	30	26	21	25	29	19	17	15	21	31	28	21	21	25	26	27
16	31	25	28	27	21	25	30	17	20	17	22	35	30	24	21	25	28	28
17	26	28	30	28	22	29	31	19	21	16	22	35	27	25	19	27	29	29
18	27	28	31	28	19	27	35	19	17	17	22	34	31	25	21	26	28	29
19	30	28	32	28	15	32	33	19	18	17	22	38	31	27	21	28	29	30
20	33	29	34	28	22	31	31	21	20	13	23	34	31	29	21	28	30	31
21	34	32	30	28	23	30	33	23	19	13	23	39	29	23	21	28	31	29
22	34	33	35	30	23	32	35	21	19	19	23	40	29	29	21	28	31	29
23	30	33	35	30	23	27	34	21	19	15	27	41	22	28	21	30	32	31
24	34	33	36	25	23	33	35	22	20	17	24	43	26	29	23	31	32	27
25	36	33	37	31	25	34	35	14	20	17	26	34	32	30	23	31	31	31
26	36	34	39	32	25	32	38	23	23	17	26	43	32	25	24	31	31	31
27	32	33	35	35	25	32	39	15	23	19	24	41	32	31	24	30	33	31
28	37	34	39	31	25	30	39	22	21	19	26	41	32	31	21	31	32	32
29	38	34	40	31	26	32	40	23	21	13	28	44	31	31	25	33	36	32
30	35	35	41	33	26	34	41	23	16	15	27	44	32	31	26	33	35	30
31	38	35	39	30	29	36	36	23	21	19	28	44	40	32	25	31	35	32
32	38	36	35	29	29	34	35	23	23	19	28	42	32	30	26	34	34	32
33	36	36	41	34	30	38	41	23	24	20	28	45	32	33	28	33	36	34
34	38	36	41	35	26	38	42	23	18	19	28	41	29	35	27	35	35	34
35	38	36	42	34	27	39	44	23	21	20	28	44	30	33	28	34	34	36
36	39	38	41	31	27	40	44	23	25	19	28	48	33	31	23	35	36	35
Mean	28.4	27.4	29.7	25.3	21.1	25.6	30.7	18.5	17.2	15.0	21.6	33.8	25.3	23.5	19.8	25.3	26.7	26.8

TABLE 5.8A.2

Covariance Matrix

(64.5)	54.6	71.2	44.2	36.7	76.5	66.5	26.3	34.4	22.4	43.5	73.3	51.0	58.8	39.4	58.6	60.2	47.3
55.5	(47.2)	61.0	37.3	31.8	64.3	56.8	23.3	28.6	18.6	37.4	62.9	40.8	49.9	33.2	49.9	51.6	39.8
71.0	61.1	(79.1)	48.7	41.0	84.1	73.7	29.8	37.6	24.4	48.5	81.5	54.6	64.9	43.4	64.8	66.9	52.0
43.2	37.2	48.5	(30.3)	25.0	52.5	45.5	17.9	23.6	15.4	29.7	50.1	35.3	40.3	27.0	40.1	41.2	32.4
37.3	31.6	41.3	25.1	(21.4)	43.2	38.2	15.7	19.2	12.4	25.2	42.3	27.3	33.5	22.3	33.5	34.7	26.8
76.3	64.5	84.0	52.3	42.2	(90.9)	78.6	30.8	41.0	26.7	51.3	86.5	61.5	69.7	46.7	69.3	71.1	56.1
65.4	56.1	74.6	46.4	38.4	78.4	(68.8)	27.7	35.2	22.8	45.2	76.0	51.2	60.6	40.5	60.5	62.4	48.6
27.0	23.3	30.3	16.5	15.9	30.6	27.3	(11.9)	13.5	8.7	18.4	30.9	18.3	24.1	16.0	24.3	25.3	19.1
34.2	28.4	36.4	24.0	19.8	41.1	34.6	13.1	(18.5)	12.1	22.9	38.6	28.2	31.3	21.0	31.0	31.8	25.3
24.5	19.5	24.0	15.4	13.3	25.8	22.8	8.0	11.9	(7.9)	14.9	25.1	18.6	20.4	13.7	20.2	20.6	16.5
43.3	37.2	48.0	30.3	25.0	51.2	45.0	17.6	24.0	14.4	(29.7)	50.0	33.0	39.7	26.5	39.7	41.0	31.8
73.4	62.6	80.6	49.8	41.6	86.8	76.3	31.6	39.7	24.3	49.9	(84.0)	55.8	66.9	44.6	66.8	69.0	53.5
51.4	40.7	55.0	34.8	27.1	61.3	51.1	19.0	28.3	18.4	32.4	55.9	(45.8)	46.2	31.3	45.4	46.0	37.8
58.8	49.5	65.4	41.2	33.6	70.1	61.1	23.3	30.8	20.8	39.7	66.4	45.4	(53.6)	35.9	53.4	54.9	43.1
39.1	33.0	43.8	27.3	23.2	47.2	40.6	16.0	20.3	13.8	26.5	43.9	31.3	35.9	(24.0)	35.7	36.7	28.9
59.1	50.6	64.1	40.4	32.5	69.4	59.8	24.1	31.9	20.5	41.6	67.4	45.3	52.3	35.3	(53.2)	54.9	42.8
60.1	51.7	67.1	40.9	35.0	71.1	62.4	25.5	31.7	19.7	40.8	69.3	46.4	54.7	36.8	54.6	(56.7)	44.0
46.7	38.9	52.8	32.5	27.2	56.1	49.2	19.6	25.1	16.3	31.6	53.0	37.8	43.6	29.0	42.1	44.1	(34.7)

Lower triangle, observed covariances
Upper triangle, reproduced covariances
Leading diagonal, communalities

TABLE 5.8A.3

Orthonormal Polynomials

1	2	3	4	5
−.2807	.3428	−.3731	.3784	−.3638
−.2647	.2840	−.2452	.1621	−.0519
−.2486	.2287	−.1360	.0031	.1314
−.2326	.1769	−.0446	−.1066	.2176
−.2165	.1285	.0302	−.1749	.2332
−.2005	.0835	.0898	−.2088	.2010
−.1845	.0420	.1350	−.2147	.1397
−.1684	.0040	.1672	−.1988	.0647
−.1524	−.0305	.1874	−.1666	−.0118
−.1363	−.0616	.1967	−.1233	−.0810
−.1203	−.0893	.1964	−.0732	−.1363
−.1042	−.1135	.1875	−.0206	−.1737
−.0882	−.1342	.1712	.0310	−.1913
−.0721	−.1515	.1486	.0787	−.1891
−.0561	−.1653	.1209	.1199	−.1685
−.0401	−.1757	.0892	.1525	−.1323
−.0240	−.1826	.0546	.1752	−.0843
−.0080	−.1861	.0184	.1867	−.0289
.0080	−.1861	−.0184	.1867	.0289
.0240	−.1826	−.0546	.1752	.0843
.0401	−.1757	−.0892	.1525	.1323
.0561	−.1653	−.1209	.1199	.1685
.0721	−.1515	−.1486	.0787	.1891
.0882	−.1342	−.1712	.0310	.1913
.1042	−.1135	−.1875	−.0206	.1737
.1203	−.0893	−.1964	−.0732	.1363
.1363	−.0616	−.1967	−.1233	.0810
.1524	−.0305	−.1874	−.1666	.0118
.1684	.0040	−.1672	−.1988	−.0647
.1845	.0420	−.1350	−.2147	−.1397
.2005	.0835	−.0898	−.2088	−.2010
.2165	.1285	−.0302	−.1749	−.2332
.2326	.1769	.0446	−.1066	−.2176
.2486	.2287	.1360	.0031	−.1314
.2647	.2840	.2452	.1621	.0519
.2807	.3428	.3731	.3784	.3638