8

Fundamentals of Common Factor Analysis

Stanley A. Mulaik

Exploratory Factor Analysis

Single Common Factor of intellectual tasks

At the turn of the twentieth century, the concepts of correlation and linear regression developed by Francis Galton, Karl Pearson, and G. Udny Yule (Galton, 1886, 1888, 1889; Pearson, 1896; Yule, 1897) were new. As psychological and educational researchers began to calculate correlations between numerous psychological and anthropometric variables (Spearman, 1897, 1904a; Yule, 1900), they became interested in why variables were correlated. Frequently this occurred in studies of intellectual abilities. Spearman (1904b) is credited with formulating the first factor analytic model to account for the correlations between six variables representing scores on five tests of academic achievement and one of pitch discrimination in 33 school children. Spearman argued that a correlation between two "series of things" (variables), if "beyond the range of chance coincidence," could be assumed "to indicate and measure something common to both series in question," (Spearman, 1904b, p. 258). The correlations between his six variables were all significantly different from zero in all cases. Furthermore the correlations were generally moderate to large. Given a collection of variables measuring intellectual activity, he held that "Whenever branches of intellectual activity are at all dissimilar, then their correlations with one another appear wholly due to their being all variously saturated with some common fundamental Function (or group of Functions)" (Spearman, 1904b, p. 273). So, the issue was how to determine if indeed a collection of variables had just a single common "factor." See the correlation matrix in Table 8.1.

Spearman concluded on the basis of procedures not regularly used in factor analysis today that his six tests had a single general factor and consequently generalized "that all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (Spearman, 1904b, p. 284).

This leads us to the model equations of his model. Let $Y_1, \cdots, Y_n$ denote n "observed" variables, like test scores, grades, trait ratings. Spearman designated the general common factor by g and by ζ_i, the *ith* specific or "unique" factor, a "specific" component

The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development, First Edition. Edited by Paul Irwing, Tom Booth, and David J. Hughes.

Table 8.1 Correlation matrix for Spearman's six variables (N = 33).

	Classics	*French*	*English*	*Mathem.*	*Discrim.*	*Music*
Classics	1.00	0.83	0.78	0.70	0.66	0.63
French	0.83	1.00	0.67	0.67	0.65	0.57
English	0.78	0.67	1.00	0.64	0.54	0.51
Mathem.	0.70	0.67	0.64	1.00	0.45	0.51
Discrim.	0.66	0.65	0.54	0.45	1.00	0.40
Music	0.63	0.57	0.51	0.51	0.40	1.00

found only in variable Υ_i. Often this is referred to as the "Two-Factor Model," which refers to two kinds of factors, found in each observed variable; the common factor and the unique factor. Then the model equations are:

$$\begin{aligned}
\Upsilon_1 &= \lambda_{1g}g + \psi_1\zeta_1\\
\Upsilon_2 &= \lambda_{2g}g + \psi_2\zeta_2\\
\Upsilon_3 &= \lambda_{3g}g + \psi_3\zeta_3\\
\Upsilon_4 &= \lambda_{4g}g + \psi_4\zeta_4\\
\Upsilon_5 &= \lambda_{5g}g + \psi_5\zeta_1\\
\Upsilon_6 &= \lambda_{6g}g + \psi_6\zeta_6
\end{aligned} \tag{8.1}$$

λ_{ig} denotes a numerical weight multiplied times the g factor variable to indicate how much a unit change of g was contributed to Υ_i. Usually g (and common and unique factors generally) are assumed to have unit variances, for convenience. Working with correlations, we also assume the variance of the observed variables is unity. Whatever was in the ***ith*** variable not due to the common factor g was the ***ith*** specific or unique factor. ψ_i denotes the numerical weight to be multiplied times ζ_i representing how much a unit change of the ***ith*** unique factor contributes to the variable Υ_i. Spearman assumed that the variables had been selected to be so diverse as to have only in common the general factor g. This assumption is not always recognized because it was thought that Spearman's model had been refuted by the discovery of other common factors. He recognized their possibility, but just wanted to focus on "general intelligence," g, and to do that, he had to select intellectual variables that did not share anything with other variables, so that whatever was not g was unique to the variable. The general factor and the specific factors are "latent" variables, meaning they are not directly observed. They are hypothetical variables invoked to account for whatever is common and whatever is unique to each observed variable. We'll deal with the question of interpreting the common factor later.

There are two major assumptions made with this model. (1) The correlations between the common factor and the unique factors are zero, $\rho(g,\zeta_i) = 0$. (2) The correlations among the pairs of unique factors are zero, $\rho(\zeta_i,\zeta_j) = 0$. This makes the general factor unrelated to any of the unique factors, and the unique factors are mutually unrelated. Generalizations of these assumptions to when there is more than one common factor, keep common factors uncorrelated with unique factors and the unique factors still mutually uncorrelated. But common factors may be correlated. Also by convention usually the common factor and unique factor variances are set to unity because they are not observed and their variances can be set arbitrarily to whatever is convenient.

Table 8.2 Factor loadings, communalities, unique variances, and unique pattern coefficients.

Variable	*λ loading*	*Communality*	*Unique variance*	*ψ loading*
Classics	.95613	.91418	.08582	.29295
French	.87078	.75827	.24173	.49166
English	.80712	.65144	.34856	.59039
Math	.74343	.55269	.44731	.66881
Discrim	.68878	.47442	.52558	.72496
Music	.65322	.42670	.57330	.75716

From the model equations in Equation 8.1 and the additional assumptions about correlations and variances between and among the common g and the n unique factors ζ_i, we can derive the correlations between distinct observed variables as

$$\rho(i,j) = \lambda_{ig}\lambda_{jg} \quad i \neq j \tag{8.2}$$

where $\rho(i,j)$ denotes the correlation between variables i and j. The correlation between variables is simply the product of the respective λ_{ig}s. The common variance in variable i, known as the ***communality***, h_i^2, of variable i, is denoted by convention as

$$h_i^2 = \lambda_{ig}^2 \tag{8.3}$$

We now need to obtain estimates for the λs and ψs in the model equation. Rather than try to do this Spearman's way, I'm going to use the modern way and obtain a maximum likelihood factor analysis solution using SPSS. I indicate that the number of observations (subjects) is 33. I set the number of factors to extract to one.

The common factor analysis yielded Table 8.2 for the values of the λs and the ψs and the communalities and unique variances.

The λ_{ig} loading is often called the factor loading or **factor-pattern** loading. It represents how much a unit change of the factor variable g will produce a change in the normalized observed variable i. In this case the loadings are substantial. The communality is the proportion of variance in the data due to the common factor. In this case, with only one factor, it is equal to the square of the factor loading of the variable. The unique variance is one minus the communality. It is the proportion of the total variance of the variable (here it is unity) due to the unique factor. And the ψ loading is the square root of the unique variance.

Assessing fit

At this point after formulating a model, collecting data, and estimating its parameters (the λ_{ig} and ψ_i), we need to determine how well the model fits the data. Here we wish to see to what extent we can accurately reproduce the correlation matrix from the estimated parameters. This we will do by comparing the original correlation matrix to a matrix of correlations reproduced by the λ_{ig} as in Equation 8.2. In Table 8.3 we show a 6 x 6 matrix. Because a correlation matrix is symmetric on either side of its principal diagonal from the left upper corner to the right lower corner, we will show the model's reproduced correlation matrix below the principal diagonal, the communalities in the

Table 8.3 Reproduced correlations below diagonal, communalities designated with* down the principal diagonal, and residuals above the diagonal.

	Classic	*French*	*English*	*Math*	*Discrim*	*Music*
Classic	.91418*	-.00258	.00829	-.0108	.00144	.00544
French	.83258	.75827*	-.0328	.02263	.05022	.00118
English	.77171	.70283	.65144*	.03996	-.0159	-.0172
Math	.71082	.64737	.60004	.55269*	-.0621	.02437
Discrim	.65856	.59978	.55593	.51206	.47442*	-.0499
Music	.62456	.56882	.52723	.48563	.44993	.42670*

principal diagonal with * to distinguish them, and the difference between the corresponding elements of the original correlations in Table 8.1 with those in Table 8.3, called the **residuals**, placed above the principal diagonal. The residuals are all near zero, which indicates good fit. But a statistical test of the fit of this one-common-factor model is given by a chi-square statistic, which has the value of 2.5157 for 9 degrees of freedom, which is **not** significant with $p = .9805$, indicating **support** for the model. The degrees of freedom are computed by the formula $df = n(n+1)/2 - m$, where df are the degrees of freedom, n the number of observed variables, $n(n+1)/2$ the number of distinct observed elements in the symmetric correlation matrix – the elements on the principal diagonal and on one side of the diagonal – and m the number of estimated parameters. Here, with n equal to 6, the number of distinct elements is 21 and the number of estimated parameters (the λ_{ig} and ψ_i) is $6 + 6 = 12$. Hence $21 - 12 = 9$. Degrees of freedom represent the number of dimensions in which the data is free to differ from the model (Mulaik, 2001).

Any possibility of difference between the reproduced and the observed correlations is due to the constraints on the model: the number of common factors, the zero correlations between and among the common and unique factors. The estimated parameters are estimated conditional on the assumed constraints of the model. Because of the small sample size of $N = 33$, there is very little power to reject the hypothesis of a single common factor. So, one needs to take this good fit with a grain of salt.

Interpretation

At this point the issue of interpretation of the results arises. This concerns relating the numerical quantities to something in the world. It must be understood that Spearman's British colleagues were operating under the influence of the philosophical school known as British empiricism: In that school all knowledge begins with sense data and gets formulated by repeated associations in space and time of similar sensory experiences. So, the researcher seeks to make **inductive generalizations** from experience. What that boiled down to was that British scientists (other than their physicists) tended to look at data and seek what was common to them. In effect their method was "exploratory." Spearman may have instead held a belief in the preexistent reality of general intelligence in the kinds of tasks that seem to call it forth and thought to demonstrate that in a confirmatory manner in his 1904 study. But exploratory factor analysis is how his British colleagues interpreted the method and began to use it to discover what is common to various intellectual performances by examining their content. Cautiously, Spearman

really doesn't come up with a definitive interpretation of *g* in this study. It is enough for him to establish that there is a single common factor among the variables he studied and a few other facts about how men and women do not differ in its capacity, how children seem to increase in their capacity until about 9 years of age, how it controls the relative position of children in school, and is to a large part responsible for success in pitch discrimination. It is important to understand that factor analysis does not yield its interpretation. Interpretation is a somewhat subjective process for the researcher to perform. It involves comparing what you have from the analysis and what you know about what occurs when a person solves problems given in tests. And the more experience you have had with things like those in the variables studied the better you will be at successfully interpreting a factor. In a way, what you are doing in interpreting a factor is formulating hypotheses that you will want to test in the future. Spearman already was testing his hypothesis.

Multiple Factors of the Intellect

Other researchers built on what Spearman had shown in his 1904 paper. They extended his methods to other tests of intellectual functioning and in doing so they discovered more than one common factor among these tests. Their reflecting on the many factors they were revealing led them to seek how to arrange them with respect to one another. By 1930 the British psychologists had developed a hierarchical theory of the intellect, with *g* at the pinnacle common to all intellective functioning, and other abilities beginning with those not as broad as *g* in their applications like verbal ability, spatial and mechanical ability, numerical ability, then descending down to more and more specific abilities confined to small groups of intellectual tasks, and finally to individual tasks (Vernon, 1961). It was like roots branching out from the trunk in large roots, to be followed by further branching into smaller roots, and eventually into small fiberous roots.

The British school of factor analysis also developed new methods for extracting many factors, which led to a growing recognition that their method was heavily based on the mathematics of linear algebra and the use of matrices. They extended the idea of an average of scores on a single variable to an average variable by adding all variables in a study together in seeking a general factor and other factors distinct from it. This came to be known as the "centroid method."

But the prominence of the British school of factor analysis was to be displaced in the 1930s and afterwards by developments in the United States. In the late 1930s L. L. Thurstone at the University of Chicago disagreed with the hierarchical structure of the intellect with general intelligence at the peak. He was suspicious of the general factor because it was usually the first factor in a centroid or principal axes solution, which he thought was just an artifact of the method of extraction. He was influenced by the theory of color vision that regarded the various hues as combinations of primary colors. Thus, by analogy, he sought to discover the "primary mental abilities" that occur in various combinations in individual tasks. He also was not satisfied with the factors produced by the summation of variables due to Burt, which came to be known as the centroid method. They simply helped us find the number of common factors by discovering at any point the additive combination of the variables that extracted the next most variance among the observed variables, after preceding factors, regarded now as additive combinations of the common parts of the variables, had been extracted. But the resulting factors were not the primary mental abilities.

There is an indeterminacy known as **rotational indeterminacy** in the common factor model. This holds that factor solutions are not unique. Different sets of factors having the same number of factors may be found that are linear transformations of one another and these could be used to reproduce the same correlation matrix. The problem was which, if any of these should be preferred? Thurstone's solution to this problem was **rotation to simple structure**. We will discuss this further later in the paper.

Thurstone's **simple structure** solutions allowed the common factors to be correlated. The factors tended to be linked to sets of quasi-homogeneous variables with large loadings on one factor and only minimal or zero loading on other factors. That made it easier to interpret the meaning of the factors by whatever was common to just those variables having high loadings on a factor. But because the common factors ended up correlated, the question arose as to what was the basis for their correlation? The obvious answer was that there were "second-order common factors" common this time to the first-order factors as opposed to individual observed variables. Here a factor analysis was performed on the correlations among the first-order factors, resulting in second-order factors. And the second-order common factors, though fewer now in number, were also correlated. So the correlations among the second-order factors were factor analyzed to find the third-order factors. This process continued to ever higher-order factors until there resulted just one factor, the *g* factor found in all the factors and variables below it, or none.

World War II, which began in 1939, led to the need by the military for tests of mental abilities that were crucial for selecting personnel with different aptitudes for training as officers, noncommision officers, and enlisted men in specialized areas. While a number of factor analytic studies were conducted during the war, they were limited by the computational burden required. After World War II, in the late 1950s computers were developed, and psychologists began to take advantage of this. Prior to computers, a factor analysis might take more than a month to complete. Today they can be calculated in seconds. Factor analysis was used to create general tests for both military and civilian use of multiple aptitudes where a number of test items, respectively, would contribute to testing each of certain intellectual abilities. Initially items thought relevant to certain intellectual abilities were assembled and subjected to factor analysis to discover the common factors that differentiated between the items. Items that did not relate with other items to a given factor were either discarded or new items were created to measure suspected new common factors.

J. P. Guilford (1967) reported on a number of studies done during World War II with the Air Force that produced 120 common factors from several studies. This led to a need to organize these factors in a meaningful manner. Guilford classified each intellectual test according to three facets. These facets were (1) mental operation (evaluation, convergent production, divergent production, memory, cognition), (2) content operated on (visual, auditory, symbolic, semantic, behavioral), and (3) the product of the operation on the content (units, classes, relations, systems, transformations, implications). Each facet contained a number of abilities, kinds of content, or kinds of products, respectively. So a distinct factor could be identified by the ability it measured, the kind of content operated on, and the result produced.

Raymond B. Cattell (1905–1998) was an English psychologist who studied under Spearman, with influences from Cyril Burt and R. A. Fisher, at the University of

London, where he obtained his Ph.D. in 1929. He emigrated to the United States in the late 1930s and after several academic appointments ended up at the University of Illinois. He conducted a profusion of factor analytic studies of the intellect, temperament, and personality. In the area of intellect, influenced by Spearman's theories of *g* and Thurstone's primary mental abilities, his major contribution, with his student John Horn (Cattell & Horn, 1978), was a factor analytic theory of intelligence with two major general factors, **fluid intelligence** and **crystalized intelligence** and a number of other factors of more restricted application. These originally were a division by Cattell of Spearman's *g* into two broad general factors. **Fluid intelligence** involves "...skills of perceiving relationships among stimulus patterns, drawing inferences from relationships, and comprehending implications" (Horn, 1988, p. 660). This factor is initially identified by a second-order factor analysis of first-order factor correlations, and it is interpreted by the nature of the test items subsumed under the common factors that have high second-order factor loadings on this factor. It has affinities to Spearman's **eduction of relations**, which Spearman defined as the power a person has when given to consider two things "to bring to mind any relations that essentially hold between them" (Spearman, 1927). It also has affinities to Louis Guttman's **analytic ability** or **rule-inferring ability** (Guttman, 1965). Both Spearman and Guttman derived these, not from factor analysis, but from examining test-item tasks as to what is shown the subject and what they require a subject to do. In items of eduction of relations or rule inferring, the subject is shown two or several pairs of things and is supposed to infer a **relation** between pairs of them. Guttman (1965) regarded intelligence as about correct perceptions of logical aspects of relations. He drew upon the concept of a relation in mathematics as a subset of a Cartesian product of two sets. The Cartesian product generates ordered pairs of elements from the two sets. A subset of these pairs is a relation between the sets. In the present case, where the subject had to demonstrate a correct perception of the relation, the subject might be required to fill in the blank for another pair in the relation where only the first member is shown. For example, "**This** is to **That**, as "**Now** is to **?**" (Answer: **Then**). Or "Dog is to puppy as Cat is to **?**." (Answer: **Kitten**, for the relation of **names of offspring of animals**). To some extent the variation in ability to answer these questions correctly depends on being able to grasp the rule of what is required by the instructions, which takes general intelligence, which may be part of the basis for correlations across different tests.

Crystallized intelligence is the general ability to correctly apply a broad range of widely applicable learned concepts, rules, terms, vocabulary, logic, arithmetic, mathematics, general knowledge, and/or practical judgment to the solution of well-defined problems. Again it is a second-order factor interpreted by what is common to tests under first-order factors that define the second-order factor. It has affinities to what Spearman called the **eduction of correlates**: "when a person has in mind any idea together with a [specified] relation, he has more or less power to bring up into the mind the correlative ideas" (Spearman, 1927, p. 166). For Guttman (1965) this is **achievement** or **rule-applying ability**. Shown a rule or given the name of its relation and a member of a pair in the relation, give the name or show correctly the other member of the pair in that relation. For example, "Who is the first President of the United States?" The relation is between the set of first presidents (of countries) and countries. Given "United States" as a member of a pair in the relation of first presidents to countries what is its counterpart? (Answer: Washington). "What is the area of a circle with a two-unit radius?" The relation shown in the instructions to the item is **areas of circles as functions of their**

radii. The subject needs to have learned that $A = \pi r^2$ and applies this rule to this problem, $A = 3.14159 \times (2)^2 = 12.5664$. So, now we can tell whether our interpretation of the g factor as **rule inferring** is correct if we can examine the instructions of all the the tasks of the variables under this factor and determine that they require rule inferring, but other tasks under Crystalized Intelligence do not involve rule inferring, but, instead, rule applying. Showing what does not indicate the factor helps as much as showing what indicates the factor. This is an objective method for verifying the objective validity of an interpretation of factors.

More recently, John B. Carroll (1993) conducted a massive survey involving factor analytic reanalyzes of 460 data sets of cognitive abilities in nine domains of cognition, from language, reasoning, memory and learning, visual perception, auditory reception, idea production, cognitive speed, knowledge and achievement, and psychomotor abilities. These analyses were conducted on the main-frame computers at the University of North Carolina in Chapel Hill. At the conclusion of his book he attempted a synthesis of the findings that resulted in a hierarchical theory of cognition in three strata. He described it as

> ...an expansion and elaboration of factor models proposed by Spearman, Holzinger, Thurstone, P. E. Vernon, R. B. Cattell, Horn, and others. It is fundamentally different from taxonomic theories such as those proposed by Guilford and Guttman, but can be accommodated within, or show correspondences with, radex theories that assume hierarchical structures. (Carroll, 1993, p. 654)

Subsequently, this has come to be known as the CHC theory, named after Cattell, Horn, and Carroll. The three strata are roughly first-order factors of test items, second-order factors of analyses of correlations among first-order factors, and the general factor g at the third stratum.

My own view on these factor analytic studies of cognition is that Carroll has brought coherence to the realm of factor analytic cognitive abilities by a uniformity of methods of factor extraction and rotation followed by what is known as a Schmid–Leiman procedure (Schmid & Leiman, 1957; see also Humphreys, Tucker, & Dachler, 1970) that orthogonalizes a set of higher-order factors into orthogonal factors at a first-order level. This makes it possible to show what degree of contribution a higher-order factor has on the individual raw variables. But I think when he called Guttman's method merely a taxonomic theory, he failed to grasp the full import of Guttman's faceted theory of intelligence as a method for external validation of interpretations of factors.

Interpretation of the factors is best when there is a close coordination between the factor analytic results and external (nonfactor analytic) analyses of test items according to classification schemas based on the task analyses of what the subject is required to do and produce. Guttman's approach to developing such schemas, called **facet analysis** (Guttman, 1994) is much more systematic and comprehensive than Guilford's. I recommend the use of facet analysis in gaining an understanding of the task requirements of items to provide a better and more objective, differentiated understanding of the nature of the processes to be associated with the various factors at different levels or strata. Combined with the Schmid–Leiman results, the facet analysis may provide a more objective and reproducible understanding of the factors. Hopefully the classification schemas may be able to reproduce the hierarchy of the common factors.

Personality

The concept of a general factor was extended beyond intelligence into the realm of personality. Webb (1915), Rushton and Irwing (2011) described a general common factor he called "will power" and designated as w. He based his analyses on ratings by peers and teachers on seven-point rating scales of 200 college students they knew well. This was the first application of factor analysis outside of the area of the intellect.

The area of personality is different from that of cognition and the intellect. Cognitive and/or intellectual abilities are studied by tests on which subjects perform. Their performance is evaluated by objective criteria of correctness given by the tester. That subjects respond similarly to different tasks may be because of the similarity those tasks have in invoking similar responses. In contrast, personality traits and characteristics are frequently judged by external raters, or by the subject of him or herself, on verbally defined scales. The judgments are formed sometimes from fleeting observations of a subject or at other times from long-time, close, intimate relations with the subject. The external perspective of others may produce different reports than from the internal perspective of the self-rater. So the accuracy of the rating as an objective description of the person can at times be questioned. One may also question whether correlations between rating scales are due to causal sources of variation within the subject, neurophysiological processes, or in external appearances of the subject to the raters that provoke similar ratings on trait scales having similar meanings.

An analogy here is with color versus hues in color vision. Color concerns the objective wavelength of light, while hue is a subjective response to light. A given hue like "yellow" may be produced by either a specific wavelength or by a mixture of two distinct wavelengths ("red" and "green"). Similarly, some traits rated may be determined by more than one distinct kind of behavior exhibited by the subject. "Friendliness" may be apparent for different reasons. In one person "friendliness" is a spontaneous trait. In another it is cultivated for deceptive purposes.

Lexical similarity between trait scales can also produce the correlations. For example to describe someone as "very friendly" is similar to saying that he is "somewhat loving." Also because many distinct traits, like "friendly," "incapable," "sociable," "domineering," "conscientious," and "self-centered" provoke implicit value judgments ("this is a good/bad person"), then a broad second-order general factor across many trait scales with diverse meanings may be an "evaluative factor," which is a subjective value judgment of the rater. Thus because some of the raters' judgments concern the conceptual dimensions on which they are able to describe or characterize persons and the artifact of similarity of lexic meaning between trait scales, these lexic rater dimensions will be confounded with dimensions on which the subjects objectively vary within themselves under the skin. In other words there is a confounding of rater dimensions with ratee dimensions (Mulaik, 1964).

Hans J. Eysenck, a German immigrant to England in the 1930s, received his Ph.D. from the University of London under Sir Cyril Burt, one of the prominent English factor analysts. Eysenck worked in hospitals during World War II and there began developing self-rating personality tests for diagnostic purposes. In 1947 his book *Dimensions of Personality* described two major dimensions, **Extraversion** and **Neuroticism** derived by factor analysis from his questionnaires. In the 1970s he added a third dimension **Psychoticism** to his theory. Eysenck interpreted extraversion as due neurophysiologically to a condition of "higher cortical arousal." Neuroticism is shown in anxiety, guilt, depression, envy, and jealousy and interpreted as a condition of a more reactive sympathetic nervous system that is more sensitive to environmental stimulation. Psychoticism is believed associated with levels of dopamine in the brain. It has been criticized as not being a unitary trait.

In the 1940s, Raymond B. Cattell also produced a famous personality inventory, the 16PF using factor analysis. The test has gone through several revisions, with the last, Fifth Edition, in 1993. It has 185 multiple-choice questions about the individual's behavior, interests, and opinions in daily life situations. It does not ask the individual to rate him or herself on certain personality traits, which some personality inventories do. Factor analysis of the questions has yielded 16 personality factors (hence "16PF").

A series of research studies by several researchers have converged to what they regard as the five major factor dimensions underlying correlations among numerous terms and phrases descriptive of personality used in self-ratings. These have become known as "The Big Five" factors, and are now in clinical use. The Big Five factors are based on self-report responses to multi-item questionnaires. The factors are (1) extraversion versus introversion, (2) agreeableness versus antagonism, (3) conscientiousness versus undirectedness, (4) neuroticism versus emotional stability, and (5) openness to experience versus not open to experience. These are each defined by about 10 items to which subjects indicate whether they agree or disagree on five-point rating scales, from Strongly Disagree through to Strongly Agree.

The author, Mulaik (1963), however, has developed a theory of personality concepts representing an expansion of Osgood, Suci, and Tannenbaum's (1957) finding of three major concepts in factor analyses of the similarity in meaning of words: **evaluation**, **potency**, and **activity**. Evaluation concerns good-bad judgments; potency judgments of power, strength, forcefulness, and bigness; active-inactive judgments of fast-slow, moving-unmoving, varying-unvarying, changing-unchanging, lively-lifeless. These same judgments occur in reactive concepts, where someone reacts to something else.

The author's theory developed a three-facet model of personality concepts based on an expansion of Osgood, Suci, and Tannenbaum's (1957) "Big Three" factors of meaning. The Big Three used here are **Good-Bad**, **Strong-Weak**, and **Active-Inactive**. Corresponding words implying a reaction to something can be formed from these, and so we have two forms of action in a facet called "Mode of Action." The Big Three are basic concepts. A third facet represents different domains of the person:

$$\underset{\text{Mode of Action}}{\left\{\begin{matrix} \textit{action} \\ \textit{reaction} \end{matrix}\right\}} \times \underset{\text{Concept}}{\left\{\begin{matrix} \textit{good-bad} \\ \textit{strong-weak} \\ \textit{active-inactive} \end{matrix}\right\}} \times \underset{\text{Domain}}{\left\{\begin{matrix} \textit{affect} \\ \textit{intellect} \\ \textit{psychomotor} \\ \textit{self-interest} \\ \textit{interpersonal} \\ \textit{work} \\ \vdots \end{matrix}\right\}}$$

The first two facets generate a Cartesian product of six basic concepts, which are **Good-Bad, Strong-Weak, Active-Inactive, Resistance to Internal Change, Resistance to External Change**, and **Evaluation with an External Standard**. A third facet is the Domain of Personality, and in Table 8.4 I have indicated six domains, but potentially more may be identified. In each domain there are six corresponding variants of the six basic concepts. These domains tend to be independent of one another. But metaphors from one domain to another are possible and make the domains not totally independent. The Cartesian product of the Cartesian product of Mode of Action and Concept with Domain of Person produces 36 distinct concepts of personality. Many of the factors found in Table 8.4 were found in other studies in the literature.

In a factor analytic study each of these potentially could correspond to a common factor, if at least three or more synonymous traits were included for each of the 36 hypothesized factors. We will illustrate farther on a factor analytic study of the six hypothesized factors in the row of the Interpersonal Domain of the table.

Factor analysis has also been used to study attitudes, organizational behavior, personality of apes and zebra finches, word meanings, beliefs, political opinions, anthropological artifacts, and nations, to name a few of many topics, and reference to these may be found in searches in the Webb using "factor analysis," and the respective topic.

Multiple Factor Analysis

We are now going to consider in greater detail the common factor model with more than one common factor. Usually the factor analytic model is expressed in matrix algebra, because this is a more compact notation, and there is a well-developed mathematical theory for this algebra. However, to make this chapter useful for those without a background in matrix algebra, we will use equations from ordinary algebra.

As in the single common factor model, we will have an algebraic equation for each observed variable. We will use Greek letters for parameters and latent variables. Capital Roman letters will stand for observed variables:

$$\begin{aligned} \Upsilon_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \cdots + \lambda_{1k}\xi_k + \cdots + \lambda_{1r}\xi_r + \psi_1\zeta_1 \\ \Upsilon_2 &= \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \cdots + \lambda_{2k}\xi_k + \cdots + \lambda_{2r}\xi_r + \psi_2\zeta_2 \\ \Upsilon_3 &= \lambda_{31}\xi_1 + \lambda_{32}\xi_2 + \cdots + \lambda_{3k}\xi_k + \cdots + \lambda_{3r}\xi_r + \psi_3\zeta_3 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \Upsilon_n &= \lambda_{n1}\xi_1 + \lambda_{n2}\xi_2 + \cdots + \lambda_{nk}\xi_k + \cdots + \lambda_{nr}\xi_r + \psi_n\zeta_n \end{aligned} \tag{8.4}$$

$\Upsilon_1, \Upsilon_2, \cdots, \Upsilon_i, \cdots \Upsilon_n$ are n observed variables. $\lambda_{i1}, \lambda_{i2}, \cdots, \lambda_{ik}, \cdots \lambda_{ir}$ (lambda) are factor-pattern loadings representing how much a unit change in a specified common factor ξ_k (xi) will produce a change in the ith observed variable. $\xi_1, \xi_2, \cdots, \xi_k, \cdots \xi_r$ denote the r common factors (variables). $\psi_1, \psi_2, \cdots, \psi_i, \cdots \psi_n$ (psi) denote n unique factor loadings (one for each unique factor variable). And $\zeta_1, \zeta_2, \cdots, \zeta_i, \cdots \zeta_n$ (zeta) denote n unique factor variables (one for each observed variable). Each equation states that an observed variable is a distinct linear (additive) combination of the r latent common factors and the latent unique factor associated with that observed variable.

Again the additional assumptions made in the single common factor model apply: The correlation $\rho(\xi_k, \zeta_i)$ between any common factor ξ_k and any unique factor variable ζ_i is zero; that is, $\rho(\xi_k, \zeta_i) = 0$. The correlation $\rho(\zeta_i, \zeta_j)$, $i \neq j$ between any two distinct unique factors is zero. However, we do not make the assumption that the correlation

Table 8.4 Mulaik's Table of Personality Dimensions.

DOMAIN	Good vs. Bad	Action: Strong vs. Weak	Action: Active vs. Inactive	Reaction: Resistance to Interbal Change	Reaction: Resistance to External Standard	Reaction: Evaluation by External Standard
Affect	Euphoria vs. Melancholy	Strength of Affect	Spontaneity of Affect	Repressed Affect	Labile Affect	Appropriate Affect
Intellect	Evaluation	Intelligence	Creativity	Memory	Field Independence	Convergent Thinking
Psychomotor	Gracefulness	Physical Strength	Impulsion	Static Poise	Physical Flexibility	Dynamic Poise
Self-Interest	Narcissistic	Domineering	Frivolous	Suppressed vs. Impulsive	Stubborn vs. Compliant	Subjective vs. Objective
Interpersonal	Positive Regard	Dominance	Sociability	Prohibitive	Nonconformity	Fair vs. Unfair
Work	Conscientious vs. Careless	Skillful vs. Inept	Ambitious vs. Lazy	Steady Worker	Persistent	Has Business Sense

$\rho(\xi_k, \xi_h)$ between any two distinct common factors is zero. It might be zero, but it also could be nonzero. But we can show mathematically that the correlation between any two observed variables Υ_i and Υ_j is due only to their loadings on the common factors and the correlations among the common factors, that is, $\rho(\Upsilon_i, \Upsilon_j) = \sum_{h=1}^{r}\sum_{k=1}^{r} \lambda_{ik}\lambda_{jh}\rho(\xi_h, \xi_k)$. The correlations between common and unique factors and between any two unique factors all being zero, the unique factor terms do not enter into this equation. When the common factors are all mutually uncorrelated, that is $\rho(\xi_h, \xi_k) = \begin{cases} 1 \text{ if } h = k \\ 0 \text{ if } h \neq k \end{cases} h, k = 1, \cdots, r$ then $\rho(\Upsilon_i, \Upsilon_j) = \sum_{k=1}^{r} \lambda_{ik}\lambda_{jk}$. In the case of uncorrelated common factors the correlations are just due to the sum of products of the factor loadings of the two variables on the r common factors. Uncorrelated factors can be mathematically convenient to work with in some cases because of this simplification. But generally, in the world, common factors can be correlated.

The common factor model separates the total variance of an observed variable into two components: common variance, or communality, and unique variance. In contrast, the theory of true and error scores separates the total variance of a variable into the sum of true and error variance. Combining these two ways of composing the total variance into a single equation, we may write

$$\text{Total variance} = \overbrace{\text{common variance} + \underbrace{\text{specific variance}}_{} }^{\text{True variance}} \underbrace{{} + \text{error variance}}_{\text{Unique variance}}$$

Unique variance consists of both specific-true variance and error variance. True variance consists of common variance plus specific-true variance. Common variance is thus always a lower bound to true variance. Dividing true variance by total variance yields reliability. Dividing common variance by total variance yields the communality of a variable. The communality of a variable represents the portion of the total variance in the observed variable due to the common factors, or common variance. Communality is a close lower bound to reliability, surpassing the squared multiple correlation for predicting a variable from all the other variables as a lower bound estimate of reliability. If the common factors are correlated, then the communality is given as $h_i^2 = \sum_{h=1}^{r}\sum_{k=1}^{r} \lambda_{ik}\lambda_{ih}\rho(\xi_h, \xi_k)$. In the uncorrelated-factors case it is given as $h_i^2 = \sum_{k=1}^{r} \lambda_{ik}^2$. Note: these formulas are exact only when the common factor model is appropriate for a given set of variables and the parameters are regarded as population parameters.

Illustration of An Exploratory Common Factor Analysis

There are a series of stages at each of which certain things must be done in conducting a factor analysis. The first stage involves choosing the variables you will analyze. Keep in mind that your data will be scores of each subject on each variable you have chosen in a large sample. That will be essential for computing correlations between each variable with every other variable in your study. Factor analysis begins with the matrix of pair-wise correlations between every distinct pair of your your variables. So, avoid whenever possible missing data where some subjects do not have scores on some of your variables.

There are work-arounds for cases of missing data, but they degrade to some extent the quality of your study. Tossing out cases having missing data loses the excluded cases' information on pairs of variables for which there is no missing data. That may work if you are sure the missing data is completely at random and in only a very small percentage of your cases. Computing correlations between each pair of variables for just cases in which there is no missing data, will make the correlation matrix be what mathematicians say is "not positive definite" because it will not be a "gramian matrix" obtained by multiplying a score matrix times its transpose, which is what is involved in getting a proper correlation matrix. So it is better to find some value to impute in place of the missing values.

Generally, the theory of imputation is that one tries to use information that is available in other variables and cases to determine the missing scores on a given variable in a given case. But there are numerous such methods and they often involve advanced statistical theory that is beyond the level of this article. So I will simply recommend that if you have missing data, do a search of the Web for "missing data" and "missing data factor analysis." Also consult the manual for the computer program you use for how it handles missing data for computing a correlation matrix and pick the method that seems most appropriate for your data.

In some cases you may simply be interested in determining the common factors among items in, say, an existing attitude inventory. So you will have little freedom in what variables you will study. You might be interested in finding the common factors among course grades and aptitude inventories available in student records. Each item in the inventory and each course grade and aptitude test score in a student's record will be a variable in your analysis.

Simply collecting an arbitrary collection of observed variables and getting measurements on numerous cases of them to compute their intercorrelations will usually not produce easily interpreted factors. The better situation to be in is where you are designing an inventory or developing a theory of the common factors in some domain of interest. There you will have the freedom to create variables you think are needed. It helps to have a clear idea of the domain of variables you wish to study and even hypotheses about what and how many common factors span this domain. If you are studying cognitive, personality, or attitude variables, then you need to review the literature to determine what factors may have been found by others and consult examples of items measuring these factors. You may need to create new items to add to those already in the literature for measuring a given factor or replace some with new ones you think better reveal the factor. Or if the factor is new, you need to work out rules for how to construct items to represent that factor and create at least three and preferably four or more of such items.

Thurstone (1937) set the standard for how to do an exploratory factor analysis. He put forth five conditions basic to common factor theory a factor analytic study should meet:

1 The number of basic factors should be less than the number of (observed) tests.
2 The diagonals of the correlation matrix (the communalities) must be regarded as unknown (and estimated).

3 The axes (factors) must be rotated into a simple (structure) configuration.
4 Each factor must be overdetermined by appearance in several (three or more) tests.
5 Tests (observed variables) should have simple factorial composition.

Let us focus here on (4) and (5). Condition (4) states that a common factor will only be determined if three or more variables in the study have strong loadings on the factor. Thurstone held that science requires evidence of objectivity for a latent construct to be obtained from different but coherent observations. So, you must already be anticipating the factors to result even as you select or create variables. And you make a special effort to include variables that measure these factors. Furthermore this is essential if you are to avoid ambiguity in interpreting the psychological content of the factor. Condition (5) requires that you include in your study variables that are relatively "pure" in measuring just one common factor, respectively, and exclude variables that measure more than one. Furthermore the assumption that unique factors are uncorrelated means that you should try not to have two indicators of a common factor that also have additional common sources of variation just between them. In my view excluding variables that depend on more than one common factor is not absolutely essential if you have at least three or four other variables that are relatively pure and strong in measuring each common factor, respectively. Simple structure will allow for variables being dependent on more than one common factor since they will fall in the hyperplanes spanned by the factors at their intersections. But in early studies it is better to have only pure representatives of the common factors. In later studies, where you are interested in how other tests depend on the basic common factors, you can include variables that depend on more than one common factor.

I will now illustrate a factor analysis using an unpublished study conducted under my supervision by Dan Griffith, for a student senior thesis at the School of Psychology at the Georgia Institute of Technology in 1998. A Power Point presentation of this study was given by Griffith and Mulaik (1998) at the Society for Multivariate Experimental Psychology (SMEP). The original study was intended to test my theory of personality trait factors described earlier in this article. It was not a complete test. We used four synonomous trait items respectively to represent each of the six potential factors, respectively, from just two personality domains – Interpersonal and Work – plus the Anxiety factor from the Affect domain. That involved 13 potential common factors. A group of 175 Georgia Tech introductory psychology students of both sexes each rated, respectively, three stereotypical persons selected at random from a large list of stereotypes on 52 bipolar seven-point trait-adjective scales, for example:

loving : 1 : 2 : 3 : 4 : 5 : 6 : 7 : unloving

That produced ratings on 525 cases. Inspection of the ratings led to the ratings of two cases being discarded because obviously the raters did not take the task seriously. In the end the number of cases analyzed was 523. We obtained 13 simple structure factors as expected, which gave tentative support for the theory.

Because of space limitations I will use 24 variables from this study representing just the six hypothetical factors from the Interpersonal domain. The correlations among these 24 variables are shown in Table 8.5.

Because we will do a **common** factor analysis, we will need to estimate communalities. This leads to the question of how to estimate them and extract the factors. Generally, you must first begin with provisional starting estimates of the communalities. From

Table 8.5 Correlations among 24 trait scales (corresponding elements above diagonal not shown).

ALLOWING																							
.12	**OUTGOING**																						
-.01	-.59	**NONTALKATIVE**																					
.63	.09	.00	**PERMISSIVE**																				
.24	.26	-.08	.22	**FAIR**																			
-.34	-.45	.35	-.30	-.47	**UNFRIENDLY**																		
.53	-.19	.24	.44	.13	-.16	**FORCELESS**																	
-.20	-.13	.20	-.14	.04	.07	-.06	**RESTRAINING**																
.47	-.21	.25	.43	.12	-.12	.72	-.07	**NONDOMINANT**															
-.33	-.32	.24	-.24	-.47	.57	-.22	.05	-.19	**UNLOVING**														
-.14	-.02	-.04	-.15	-.04	.14	-.17	-.01	-.13	.08	**NONCONFORMING**													
.13	.07	.02	.14	.23	-.20	.15	.01	.15	-.20	-.34	**COMPLIANT**												
.17	-.07	.11	.10	.24	-.12	.15	.04	.18	-.09	-.01	.18	**IMPARTIAL**											
-.24	.36	-.31	-.17	.20	-.15	-.46	.08	-.51	-.14	.09	.04	-.06	**LEADER**										
-.37	-.38	.25	-.31	-.52	.59	-.21	.08	-.23	.67	.11	-.27	-.16	-.16	**UNKIND**									
.04	-.07	-.01	.01	-.22	.09	-.01	-.10	-.04	.12	.38	-.30	-.04	-.13	.21	**UNCONVENTIONAL**								
-.38	.26	-.19	-.29	.06	.00	-.56	.19	-.58	.04	.14	-.03	-.11	.64	.06	-.05	**DIRECTING**							
-.40	-.09	.06	-.32	-.14	.24	-.32	.21	-.26	.22	-.01	-.02	-.17	.14	.24	-.08	.22	**PROHIBITIVE**						
.02	-.27	.24	.05	-.04	.18	.13	.11	.15	.11	.01	-.01	.15	-.21	.13	.10	-.15	.03	**UNGREGAIOUS**					
.22	.26	-.09	.23	.69	-.45	.17	.07	.14	-.45	-.11	.30	.24	.21	-.59	-.27	.02	-.16	-.15	**JUST**				
.29	.00	.07	.28	.34	-.23	.28	.07	.28	-.29	-.14	.24	.32	-.04	-.34	-.15	-.20	-.22	.03	.37	**NEUTRAL**			
.17	.58	-.47	.16	.30	-.50	-.06	-.08	-.05	-.37	-.05	.13	.00	.34	-.40	-.15	.21	-.11	-.26	.31	.15	**SOCIABLE**		
-.33	-.36	.25	-.28	-.40	.51	-.20	.09	-.22	.62	.07	-.20	-.12	-.09	.59	.11	.04	.23	.13	-.41	-.27	-.46	**UNAFFECTIONATE**	
-.11	-.04	-.11	-.08	-.31	.24	-.18	-.19	-.18	.30	.31	-.38	-.12	-.06	.34	.47	.07	.04	.00	-.33	-.28	-.11	.25	**REBELLIOUS**

those you will determine the number of common factors to retain, and the iterative analysis will obtain a solution for the common factor model, including new estimates of communalities based on that number. There are many methods of factor extraction. The SPSS FACTOR program I use lists seven extraction methods: **PC** principal components (not a true common factor analysis); **PAF** principal axes factoring (least squares with squared multiple correlations for predicting each variable from the rest as initial communality estimates); **ML** (maximum likelihood); **ALPHA** (alpha factoring); **IMAGE** (image analysis); **ULS** (unweighted least squares); and **GLS** (generalized least squares). A detailed description of the mathematical theory behind these methods is given in Mulaik (2010). I will use the **ML** estimation (Jöreskog, 1967). This method usually assumes the observed variables are distributed according to a multivariate normal distribution with means of zero and variances of unity. Obviously the rating scales are not continuous variables, being "polytomous" seven-point scales. Still the ML method is robust and may be justified here by the equivalence of the equations for this model with another model that makes no distributional assumptions. Howe (1953) developed a model that sought to maximize the determinant of the matrix of partial covariances among the residual variables after r common factors were partialled from the original covariance matrix. (A matrix is a rectangular array of numbers arranged in rows and columns.) This determinant tends to a maximum when the matrix is diagonal and is near a maximum when off-diagonals are near zero. This is an algebraic criterion and not a statistical criterion. Optimizing this criterion produces the same equations to optimize as in ML, and hence produces the same solution as the ML method (see Mulaik, 1972, pp. 169–172).

ML estimation involves an iterative procedure. It usually begins with estimates of the squared multiple correlations $\hat{R}_i^2$ for predicting each variable from the rest of the variables in the correlation matrix as provisional initial estimates of the communalities. (These squared multiple correlation coefficients are given by a matrix equation we do not need to discuss here.) That is, $\hat{h}_i^2 = \hat{R}_i^2$. The squared multiple correlation is a strong lower bound to the communality of a variable (Guttman, 1956). The procedure then subtracts each of these squared multiple correlations $\hat{R}_i^2$ from 1, $\hat{\psi}_i^2 = 1 - \hat{R}_i^2$, to get the initial estimate of the unique variance $\hat{\psi}_i^2$ of the ith variable. Next it takes the reciprocals of the square roots of the unique variances $\hat{\psi}_i^{-1}$ of each observed variable and multiplies every element in the ith row and column of the correlation matrix by the respective $\hat{\psi}_i^{-1}$. Now, the rows and columns of the correlation matrix represent variables. So, these coefficients act as weights that have the effect of converting the correlation matrix to a variance/covariance matrix. Variables having small unique variances (hence large communalities) thus have larger variances, giving greater weight in the analysis to variables with high communalities.

Next the computer program computes the eigenvectors and eigenvalues of the resulting weighted correlation matrix. The eigenvectors are the columns of a matrix known as the **eigenvector matrix**. The eigenvalues constitute the elements of the principal diagonal of a $n \times n$ square diagonal matrix with zero off-diagonals, known as the **eigenvalue matrix**. The process of how this computation is done is too technical to discuss here. But there are n eigenvectors eigenvalues that can be obtained from any square symmetric matrix, like a correlation matrix. Each **eigenvector** is a column array of n numbers, $\mathbf{a}_j = [a_{1j}, a_{2j}, \cdots, a_{nj}]$, $j = 1, \cdots, n$ represented in text by elements within brackets. The sum of the squares of these numbers in each eigenvector add up to unity, $\sum_{i=1}^{n} a_{ik}^2 = 1$.

The sum of cross products of corresponding elements in two distinct eigenvectors k and h is equal to zero. That is, $\sum_{i=1}^{n} a_{ik}a_{ih} = 0,\ k \neq h$. Thus we say that two distinct eigenvectors are mutually **orthogonal**.

The coefficients in the eigenvector arrays will be used as weights to get linear (additive) combinations $X_k = \sum_{i=1}^{n} a_{ik}\tilde{Y}_i$ of the observed variables after the unique factor parts of them have been partialled out, that is, $\tilde{Y}_i = Y_i - \psi_i\zeta_i$.(Actually we never know what the unique factor scores are; but to aid understanding we act as if we do.) Here, X_k is derived by using the coefficients from the kth eigenvector as weights in adding together the n modified observed variables $\tilde{Y}_i$. So there can be potentially n such linear combinations, each a new variable. The covariance between any two linear combinations so obtained is $\sigma(X_k, X_h) = 0,\ k \neq h$. The variance $\gamma_k = \sigma_k^2(X_k)$ of the variable that is the linear combination of the observed variables using coefficients of the kth eigenvector as weights respectively to additively combine the respective observed variables is known as the kth **eigenvalue**.

The column eigenvectors are by convention ordered in the eigenvector matrix in descending order of magnitude of their corresponding eigenvalues. The eigenvalues on the principal diagonal of the eigenvalue matrix begin with the largest in the upper left corner, and procede in descending magnitude down the principal diagonal to the smallest in the lower right corner. So, assuming we have ordered the eigenvectors according to the descending order of their eigenvalues, the first linear combination X_1 has variance $\gamma_1 = \sigma_1^2(X_1)$, which is the maximum variance of any of the linear combinations. So, X_1 determines a dimension in which there is the greatest common variance in common factor space. X_2 determines a dimension orthogonal to (independent of, uncorrelated with) X_1 that has the next largest variance in common factor space. Each successive linear combination X_k determines a dimension orthogonal to the preceding dimensions with the next largest variance. At some point, if the modified observed variables $\tilde{Y}_i$ contain only effects of the common factors, the remaining linear combinations will have zero variances or zero eigenvalues.

The magnitudes and number of positive eigenvalues determines the number of common factors to retain. That's the theory anyway. In practice, with sampling error, population zero variances become nonzero, though small. Usually a number of the smaller positive eigenvalues are still near zero, while with random overestimates of the unique factor variances the smaller eigenvalues frequently are negative, and these might be joined with the zero eigenvalues to be discarded because, usually, they correspond to unanticipated minor sources of covariance between the variables that are uninterpretable. They may even correspond to residual covariation between pairs only of variables, which are not uniquely determined. The common factor model does not prescribe anything about covariation due only to two variables, since such factors in principle could equal $n(n-1)/2$ in number, which is greater than the number of observed variables and unique factors. These are known as **doublet factors** in the literature. Thurstone (1947, p. 442) thought these were produced by two variables measuring the same performance and one should be discarded because they were uninterpretable together in any unique way. Attempts to identify and eliminate doublets from the analysis are described in Mulaik (2010, pp. 253–262). Image analysis explicitly models doublets in an approximation to common factor analysis and could be used.

In practice the eigenvectors and eigenvalues are calculated directly from the correlation matrix among the observed variables with communality estimates replacing 1 s in the principal diagonal, beginning with lower bound estimates like the squared multiple correlations. Often the correlation matrix is rescaled by multiplying respective rows and columns by the reciprocals, respectively, of the square roots of the corresponding initial estimates of unique factor variances, given by $\hat{\psi}_i^{-1} = 1/\sqrt{1-R_i^2}$, where R_i^2 is the squared multiple correlation for predicting the ith observed variable from the $n-1$ other observed variables. An initial iteration is used to determine the number of factors to retain and this number made a fixed value. Then subsequent iterations are used iteratively to recalculate improved values until some criterion of good fit is achieved.

How many factors?

If one knew exactly the true values for the communalities and placed these in their corresponding places in the principal diagonal of the correlation matrix, then the number of common factors would equal the number of positive, nonzero eigenvalues of this "reduced" correlation matrix. There are a number of heuristic criteria for the number of factors to retain based on the magnitude of the eigenvalues. A common but criticized criterion is to save the factors with eigenvalues greater than 1.00 of the unmodified correlation matrix. This happens to be the weakest lowest bound for the number of common factors according to Guttman (1956), corresponding to the number of positive eigenvalues of the correlation matrix with 0 s replacing the 1 s in the principal diagonal of the raw correlation matrix. Zero is the worst possible estimate for a communality. So, this criterion may discard valid common factors with small nonzero eigenvalues. Even Henry Kaiser (1960, 1961) who first proposed this criterion recognized it as the weakest lower bound to the number of common factors (personal communication).

Another popular criterion is to plot the magnitudes of the eigenvalues, arranged in descending magnitude, against the ordinal numbers of the eigenvalues. (This should be done with eigenvalues of either the reduced or the weighted correlation matrix $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$ (see Appendix I) and not the raw correlation matrix. Unfortunately, SPSS only plots the eigenvalues of the raw correlation matrix, which is not appropriate for ML.) If you connect the dots in the plot, the resulting curve through the dots begins with a very large eigenvalue followed by rapidly descending values as the curve advances. At some point the curve may suddenly change from a steep downward slope to a more gentle straight-line drop of the curve with subsequent eigenvalues. R. B. Cattell called this the "scree point," because it reminded him of the place at the base of a steep cliff where falling rocks have rolled away in a gentle straight-line descent from the base of the cliff. Sometimes there is no clear sharp break in the slope, but a gradual curve, and this criterion is ambiguous.

Another related method the author uses is the "ruler method" (Mulaik, 2010, p. 187; Figure 8.1). This is inspired by the observation that if samples are obtained from multivariate normal populations in which the variables are all uncorrelated, the sample eigenvalues, when plotted against their ordinal value, fall on a gently-downward-sloping straight line from left to right. By symmetry, the population line passes through the average eigenvalue of $1 = n/n$ at $m = n/2$. The slope seems to be a function of sample size, the number of variables, and the multivariate normal distribution. The smaller the

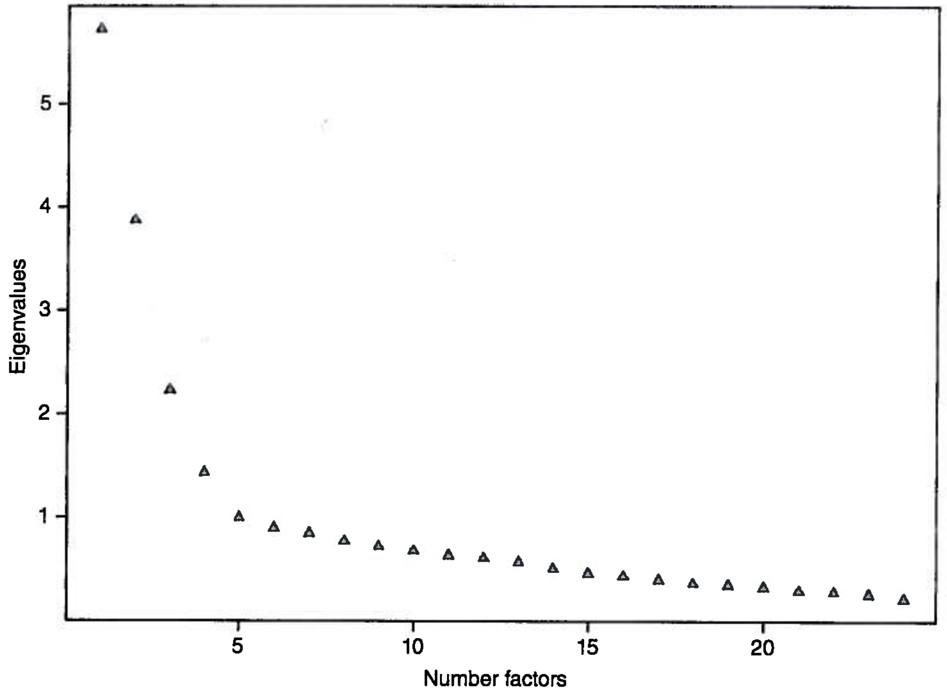

Figure 8.1 Plot of eigenvalues of matrix $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$.

sample the steeper the slope. The sum of the eigenvalues always equals the sum of the diagonal elements of the correlation matrix. In this case the average eigenvalue is $1 = n/n$. Presuming that the first r population eigenvalues of a correlation matrix with communalities in the principal diagonal should be nonzero and represent common variance, and the remainder all be zero, then after the number r of factors has been identified, the remaining sample eigenvalues should fall on a gentle downward sloping straight line to the right. So, if we fit a straight line through as many of the last eigenvalues on the right as we can, and we extend the line to the left, any eigenvalue clearly above the line should represent common variance. What is "clearly" above the line will be a subjective choice. This is an eye-ball heuristic.

A related method to the ruler method is the Parallel Analysis (PA) method for determining the number of common factors (Horn, 1965). Researchers have assembled tables of the averaged eigenvalues for numerous samples of specified size N from sample correlation matrices sampled from a population of uncorrelated variables of size n in number. If you have plotted the eigenvalues from your sample correlation matrix with communalities in the principal diagonal, you next superimpose on the plot values of the averaged eigenvalues from the tables for a similar size sample and the same number of observed variables. Where the lines cross is the number of factors to be retained. But the method seems to be biased to take too few factors. Turner (1998) noted that a very large size for the first eigenvalue of the real data would force the remaining eigenvalues to sum to a smaller number because the sum of the eigenvalues always equals the sum of the diagonal elements respectively of the correlation or the modified correlation matrix. That would lower the last several eigenvalues. Perhaps it would be better to note the

difference between the smallest eigenvalues of the real data and that of the tables, which would correspond to zero eigenvalues in the real population. The curves through the smallest eigenvalues in both plots seem to be almost parallel, straight lines. The slopes seem to be a function of sample size. The difference in level between them seems to be due to the common factor variance in the real data curve If this difference is subtracted from each of the table eigenvalues and these points plotted, you get effectively the same line as the ruler method.

There is also a series of **chi-square tests** that you can perform to test whether the number of factors is as specified. This involves a series of analyses done for each hypothesized number for the common factors. You can begin with the null hypothesis that the number of factors is one and test that. If the chi-square test yields a significant chi-square, you reject that hypothesis and go to the next, there are two common factors and reanalyze accordingly. You keep going with these tests until you just accept the null hypothesis that the number of common factors is whatever number you have arrived at.

Experience with the chi-square tests is that usually they reject the null hypothesis for any but a quite large number of common factors. And the larger the sample, the more factors must be taken until chi-square is not significant. When applied to polytomous data like the seven-point rating scale data of the present study, in a large sample the test may reject the model because the data is not distributed as a continuous multivariate normal distribution with that number of factors. Other possibilities of failure are the presence of other common factors or even doublet factors not anticipated and controlled for in variable selection and/or construction. Most exploratory factor analysts prefer the heuristic tests to the chi-square test.

Nevertheless for the current factor analysis of the 24 variables with six factors retained, the chi-square statistic was 275.5282 with 147 degrees of freedom and a p value less than .001, which is significant. This means that the exploratory common factor model with six factors did not fit the data to within sampling error. But remember that the sample size is 523, which means that the power to detect small deviations beyond chance is quite high. With large samples, degree of approximation becomes more important in exploratory studies.

There are other indices of degree of approximation of the model to the data, such as the CFI (Bentler, 1990; McDonald, 1990) and the RMSEA (Steiger & Lind, 1980), which are built from the chi-square statistic values. Mulaik (2009) discusses these and several other approximate fit indices. These have been developed in structural equation modelling and current commercial computer programs should be including these indices in current upgrades.

Simple structure rotation

At this point in the analysis the program will have computed the "unrotated factor matrix," which is formed from the first r eigenvectors, with elements in each column respectively multiplied by the square root of the eigenvalues. The correlations between the corresponding distinct factors are necessarily zero. Because ML estimation analyzes the data after rescaling the data by multiplying elements in each row and corresponding column of the correlation matrix by the reciprocals of the respective unique factor loadings $\hat{\psi}_i^{-1}$, the elements in each row of the modified eigenvector matrix is rescaled to its original metric by multiplying each element in a row by $\hat{\psi}$. I will not show this here because of space limitations, and it is not the final result we want. We may note, however, that the first column of the unrotated matrix usually has mostly larger loadings. From early on the first factor was often confused with the general factor, because almost every variable had a large loading on the first factor. But this is an artifact of the method. The general factor has to be a higher-order factor obtained by factor analyzing correlations among simple structure factors. But the unrotated factor matrix is the basis for the next step, which is rotating this initial factor loading matrix to a factor-pattern matrix having **oblique simple structure**. Technically, this involves what mathematicians call a **linear transformation** of the unrotated factor matrix.

The idea for simple structure was first proposed in the late 1930s and described in detail later by L. L. Thurstone (1947). Until computers were applied to this problem, rotations were conducted graphically, by hand, by rotating two factors at a time. It took well over a month to rotate a factor matrix by hand to oblique simple structure. We will not discuss the details of the procedures but will describe the why and the what for simple structure. These details are discussed in Mulaik (2010). But the concept of simple structure is still not well understood and many misconceptions about it still abound in the literature.

The problem is that the unrotated factor matrix is not unique. In fact, there are an unlimited number of distinct linear transformations or rotations of the unrotated factor matrix that might be obtained. All of them will generate the same reproduced correlation matrix. So, which rotation should we obtain? Thurstone saw the problem as finding an objective solution because he believed the common factors were real and should be found with different sets of variables from the same factor domain (i.e., variables that were all linear combinations of the same common factors).

Thurstone (1947) devised an ingenous method for finding the simple structure factors and their loadings. He first regarded that all variables could be represented by vectors radiating from a common origin in what he called common factor space (see Figure 8.2). Vectors in this case are just arrows. The cosines of angles between vectors would equal correlations between the corresponding variables, which constrains their positioning with respect to one another in multidimensional space. There would be arrows for observed variables, arrows for the common factor variables and arrows for what he called the "reference structure axes." The reference structure axes were to be used to determine planes or "hyperplanes" (a subspace of $r-1$ dimensions in r-dimensional space). A hyperplane would contain all variables that were linear combinations of at most $r-1$ of the common factors in those hyperplanes. And each of the r common factors would be common to $r-1$ hyperplanes, so they would be at the intersections of r hyperplanes.

So, how would we discover these hyperplanes? The reference structure axes were the solution. They would be vectors at right angles to every vector in a given hyperplane. So, if you plotted, say, two factors all the variables in the plane passing through these two factors would be orthogonal to a reference axis at right angles to these two factors, and to every variable that was a linear combination of just these two factors. We can see that in Figure 8.2, which shows a three-factor simple structure solution in three dimensions. A hyperplane in this figure is a subspace of $3 - 1 = 2$ dimensions, a plane, spanned by the two common factors. Every other variable that is just a linear combination of these two factors would fall in the plane common to the two factors. To find this plane we would use the reference axes like handles rigidly attached at right angles to the plane of two

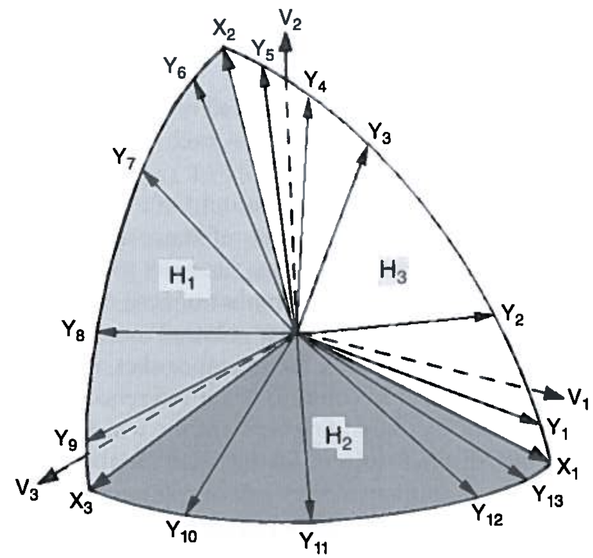

Figure 8.2 A simple structure factor solution in three dimensions showing the three hyperplanes H_1, H_2, and H_3 at the intersection of which are the factors X_1, X_2, and X_3. The reference axes are the vectors V_1, V_2, and V_3, orthogonal to their respective hyperplanes. Y-vectors are observed variables in the hyperplanes that are linear combinations of pairs of factors.

dimensions, and we would move it around in three-dimensional space and in doing so move the plane around in space until we had identified a group of observed variables that were at right angles to the reference axis in the hyperplane. We would do the same thing for each reference axis, making sure it identified distinct sets of observed variables that were at right angles to the respective reference axis and containing very few variables from any other hyperplanes. Once the hyperplanes are found, we can use the reference axes to determine the common factors at the intersections of these hyperplanes.

This may be difficult to grasp at first, but the important point made by Thurstone (1947) was that the determination of the simple structure common factors did not depend on the particular set of observed variables in the analysis. A common factor exists independently in the world and more than one set of variables is dependent on it. As long as we can select several variables for study that are linear combinations of less than all of the same factor variables, these will fall in the same hyperplanes and determine the same factors at their intersections. He also showed that mathematically you could determine the factor-pattern loadings on the common factors using the correlations or cosines of the observed variables with the reference axes.

This approach worked well with graphical rotation, but it was slow going to do it. The arrival of computers made it possible to consider analytic methods of rotation, but the problem was how to represent what a simple structure solution was like in a form a computer program could use to do the rotations. First it was essential to find a **reference structure matrix** representing correlations of observed variables with the hypothetical reference axes. By identifying variables that had zero correlations with the provisional reference axes, we could determine the reference axes. These zeros would be located in key places in the resulting reference structure matrix. Thurstone (1947) developed five rules of what the reference structure matrix would look like once we had a simple structure solution (I paraphrase):

1 So that each variable will not be dependent on all r of the common factors it must be in a subspace of at most $r - 1$ dimensions, so, each row of the reference structure matrix should have at least one zero in it.
2 To overdetermine the reference axes each column of the reference structure matrix should have r zeros with the corresponding reference axis.
3 To ensure distinctness of the reference axes, for every pair of columns of the reference structure matrix there should be a number of zero entries in one column corresponding to nonzero entries in the other column.
4 To further separate and overdetermine the reference axes, when there are more than four common factors, each pair of columns of the reference structure matrix should have a large proportion of corresponding zero entries, which forces separation of the observed variables into distinct clusters. So that each variable depends on, at most, a small number of common factors; in every pair of columns of the reference structure matrix there should be only a small number of corresponding nonzero entries. Once a reference structure matrix was found with these properties, the factor-pattern matrix could be immediately derived by a specific linear transformation.

Some of the first successful computer programs for rotating factors found "simple structure solutions" with orthogonal factors. The most popular of these is Kaiser's (1958) VARIMAX method. Generally, however, I do not recommend orthogonal factor solutions in common factor analysis. Orthogonality, or factors that are mutually uncorrelated, imposed a priori by the researcher, is an artifact and does not represent a condition in the world. So, calling an orthogonal solution a simple structure solution is a contradiction. To identify the hyperplanes and the factors at their intersection we must allow for correlated factors.

Of many programs that were developed to obtain correlated or "oblique" simple structure solutions, the most successful, in my opinion, until very recently was Jennrich and Sampson's (1966) **direct oblimin method**. This method originally rotated two factors at a time (known as **planar rotations**). They realized further that to each zero entry in the reference structure matrix there would be a corresponding zero in the factor-pattern matrix. So, rather than using a reference structure matrix as an intermediate solution, they made transformations directly to produce a factor-pattern matrix by producing zeros (or near zeros) appropriately in the matrix. The factor-pattern matrix contains the coefficients indicating how much a unit change of a common factor produces a change in an observed variable.

Jennrich (2002) made furthermore a major breakthrough in factor rotation methods by the use of an easily implemented algorithm known as the **gradient projection algorithm**. This method further transforms the whole factor-pattern matrix at a time rather than using planar rotations. That has made it possible to provide programs for numerous rotation algorithms, including some newer ones that slightly improve over the direct oblimin method in yielding simple structure solutions (Bernaards & Jennrich, 2005). Jennrich streamlines the criterion by seeking to minimize

$$\text{Quartimin} = \sum_{j=1}^{r}\sum_{k=1}^{r}\sum_{i=1}^{n}\lambda_{ij}^2\lambda_{ik}^2/4 \; j \neq k \tag{8.5}$$

This criterion is minimized when squared loadings in one column correspond to zeros or near-zero loadings in the other, and when some columns have corresponding zeros or near zeros in both columns.

Among the new methods are Geomin (Yates, 1987), Infomax (McKeon, 1968; Browne, 2001), Promin (Lorenzo-Seva, 1999), Simplimax (Kiers, 1994), several based on **component loss functions** (CLF) like Katz and Rolf's (1974) Functionplane, Rozeboom's (1991a, 1991b) HYBALL, and several simple CLF algorithms by Jennrich (2006), and Absolmin and Epsolmin (Mulaik, 2010).

Absolmin is a simple CLF criterion recommended by Jenrich and named by Mulaik (2010), which contains details of these newer methods based on the gradient projection algorithm. A component loss function seeks to minimize each element in the pattern matrix irrespective of other elements, under constraints of a linear transformation.

$$\text{Absolmin} = \sum_{i=1}^{n}\sum_{j=1}^{r}|\lambda_{ij}| \tag{8.6}$$

This simple CLF criterion seeks to minimize the sum of the absolute values of the elements of the rotated matrix. Unfortunately the solution has many local minima, and numerous randomly generated orthogonal transformation matrices should be used as starting values, and the converged solution from these that has the smallest Absolmin criterion value should be retained.

Also accompanying the factor-pattern matrix, there should be a matrix of correlations among the factors. This suggests the possible existence of "higher-order" factors common to the "first-order" factors. To save space, I have not computed a factor analysis of these correlations. Jennrich has made his programs available in several programming languages for the gradient projection algorithm (gpa) at www.stat.ucla.edu/research/gpa/.

In Table 8.6, I show the factor-pattern matrices and correlations among factors respectively for the Direct Oblimin solution and the Absolmin solution. These are the principal data for interpretation of the factors. Both solutions are quite similar. I have placed in **bold type** the largest elements, and with few exceptions they correspond to the first decimal place between the two solutions. The remaining loadings are zero or near zero. Furthermore, the large loadings determine factors that correspond to factors expected under the hypothesized model for the interpersonal domain. Factor 1 can be called **Permissiveness**, Factor 2 **Introversion**, Factor 3 **Unjustness**, Factor 4 **Negative regard for others**, Factor 5 **Nondirecting**, and Factor 6 **Nonconformity**.

One thing you may notice is that in columns for Factor 3, the loadings for the Direct Oblimin solution are generally negative of those, respectively, in the Absolmin solution. Furthermore Factor 5 in both solutions is predominantly negative, when we expected positive loadings. This occurs commonly in factor rotation solutions and is an artifact of the rotation algorithm. What this suggests is that the solution for the third factor under Direct Oblimin is the opposite of that under Absolmin. But both identify nearly the same dimension; one solution for the factor just points along the same line but in the opposite direction from the other. It is easy to make everything the same or predominantly positive by simply reversing the signs in the column in question and making corresponding adjustments for the signs of the correlations among the factors. Or you can interpret the factor according to the predominant negative loadings, if they occur.

Table 8.6 Simple structure factor-pattern matrices and correlations among factors for Direct Oblimin and Absolmin.

	Simple Structure Factor-Pattern Matrices											
	Direct oblimin						*Absolmin*					
Variable	*1*	*2*	*3*	*4*	*5*	*6*	*1*	*2*	*3*	*4*	*5*	*6*
ALLOW	**.746**	.000	-.007	.099	.060	-.024	**.727**	-.021	-.049	-.039	**.244**	.047
OUTGO	.103	**-.649**	-.045	.126	-.211	-.045	.174	**-.600**	.079	-.061	**-.263**	-.070
NONTALK	.067	**.692**	-.079	-.158	.048	-.050	.007	**.684**	.019	.084	**.167**	-.021
PERMISS	**.773**	.009	-.061	-.046	-.054	-.063	**.756**	.019	.000	.097	**.154**	.000
FAIR	.022	-.079	**-.675**	.144	-.098	.005	.090	.051	**.604**	**-.247**	-.032	.000
UNFRIEND	-.142	**.289**	.138	**-.450**	.030	.047	**-.201**	**.275**	-.127	**.423**	.040	.038
FORCELSS	**.323**	**.200**	-.063	.155	**.473**	-.092	**.322**	.147	-.009	-.170	**.579**	-.019
RESTRAIN	**-.128**	**.319**	-.079	.016	**-.206**	-.105	-.133	**.347**	.064	-.061	**-.191**	-.120
NONDOMIN	**.235**	.183	-.070	.166	**.547**	-.102	**.242**	.126	.000	**-.191**	**.629**	-.032
UNLOVING	.030	-.026	-.001	**-.875**	-.010	-.035	.000	.028	.000	**.847**	.052	-.048
NONCNFRM	-.112	.067	-.107	.082	-.089	**.632**	-.192	.059	.143	-.096	-.117	**.615**
COMPLINT	.068	.021	-.105	.057	-.022	**-.458**	.139	.060	.058	-.079	.017	**-.449**
IMPARTL	.079	.119	**-.334**	-.026	.047	.022	.090	.171	**.281**	-.039	.122	.036
LEADER	.043	-.086	-.060	.068	**-.770**	-.002	.051	.000	.096	-.038	**-.768**	-.057
UNKIND	-.037	.138	**.247**	**-.582**	-.029	.064	-.103	.121	**-.217**	**.594**	-.015	.053
UNCNVENT	.137	.060	.098	.039	.021	**.670**	.024	.000	-.057	.000	.032	**.681**
DIRECTNG	.028	.043	.059	-.014	**-.850**	.059	.000	.112	-.008	.055	**-.848**	.000
PROHIBTV	**-.335**	.062	.083	-.111	-.120	-.113	**-.325**	.073	-.055	.091	**-.198**	-.150
UNGREGAR	.070	**.348**	.061	-.005	.061	.055	.022	**.312**	-.079	-.009	.111	.076
JUST	-.070	**-.191**	**-.851**	.031	.011	-.063	.033	-.023	**.762**	-.174	.071	-.073
NEUTRAL	.161	.105	**-.298**	.128	.068	-.130	**.195**	.146	**.234**	-.178	.151	-.105
SOCIABLE	.185	**-.465**	-.021	.222	**-.234**	-.111	**.246**	**-.429**	.038	-.159	**-.252**	-.122
UNAFFECT	-.080	.108	.021	**-.651**	-.015	.015	-.118	.140	-.017	**.617**	.017	.000
REBELLOS	.084	-.186	.034	-.301	-.002	**.567**	.002	-.196	.012	.331	.000	**.559**

	Factor Correlations											
	Direct oblimin						*Absolmin*					
	1	*2*	*3*	*4*	*5*	*6*	*1*	*2*	*3*	*4*	*5*	*6*
Permissiveness 1	**1.00**	-.093	-.291	.388	.491	-.054	**1.00**	-.169	.308	-.530	.200	-.046
Introversion 2	-.093	**1.00**	-.048	-.314	.280	-.051	-.169	**1.00**	-.055	.241	.281	-.059
Unjustness 3	-.291	-.048	**1.00**	-.549	.050	.334	.308	-.055	**1.00**	-.436	-.204	-.297
Negative Regard 4	.388	-.314	-.549	**1.00**	-.028	-.270	-.530	.241	-.436	**1.00**	-.011	.243
Nondirecting 5	.491	.280	.050	-.028	**1.00**	.025	.200	.281	-.204	-.011	**1.00**	.021
Nonconformity 6	-.054	-.051	.334	-.270	.025	**1.00**	-.046	-.059	-.297	.243	.021	**1.00**

Appendix II contains the MATRIX – END MATRIX program from SPSS to obtain the Absolmin solutions. The program requires that you begin with an orthogonal solution for a factor-pattern matrix. It could be the initial unrotated principal axes factor-pattern matrix, or it could be an orthogonal rotation of that matrix. In this case we begin with the VARIMAX solution. The iterations begin with a random orthogonal transformation matrix and the gradient projection algorithm is used to find a solution that minimizes the absolute value of each element of the rotated factor-pattern matrix. The Absolmin algorithm yields numerous "local minima." So the way to use it is to begin the iterations many (e.g., 100) times with different random transformation matrices and compare the criterion values of each solution. Pick the solution with the smallest criterion value.

Factor scores and factor indeterminacy

Having examined the factor-pattern and factor-correlation matrix to determine interpretations for the factors, the next question that may arise is "How can we get scores on the common factors?" My personal preference is to avoid factor scores unless the squared multiple correlation for predicting the common factors is in the high .90s. My reason for this opinion is **factor indeterminacy**. Factor scores are generally obtained by regression. Regression estimates give the least squares estimates of the factor scores by predicting them from the observed variables using the correlations among the observed variables and the correlations of the variables with the factors given in what is known as the factor structure matrix to compute the regression weights for the prediction of the scores. Other methods cannot get better least-squares fit than the regression method. The concept of factor indeterminacy has received an extensive treatment in the literature, and I and R. P. McDonald have been major contributors (Mulaik, 1972, 1996a,b, 2005; Mulaik & McDonald, 1978). But briefly the problem is this: The common factor model has n observed variables and $n+r$ latent common and unique factor variables. The observed variables are thus determined by what is not observable. There is no way to turn around to get unique determinations of the more numerous common and unique factors from the smaller number that is observed. We can get the factor structure matrix of correlations between the observed variables and the common factors, for these are derived from the factor-pattern matrix multiplied times the factor correlation matrix, which are determinable. And the correlation matrix among the observed variables is observed (we based our factor analysis on it). So, we can get regression weights for predicting the common factors from the observed variables. But these predictions frequently have multiple correlations less than .7071. Now, let me illustrate a simple analogy.

Suppose I have a variable X and I say that it is correlated .7071 with some unknown variable Y. Can you tell me from this what variable Y is? You are smart and say "No!" Why? Because, you say, there are an unlimited number of other variables that might be correlated .7071 with X. But, I ask, haven't we narrowed down the problem of what Y is from knowing it is correlated .7071 with X? "Well, yes," you say. "Suppose we represent variables by vectors emanating from a common origin, with cosines of the angles between them positioning them with respect to one another in space. Correlation = cosine of angle. Then the set of all variables Y having the correlation of .7071 correspond to vectors in a cone surrounding X as the central axis of the cone with cosine of .7071 for the angle between the X vector and any vector Y in the cone." The number of vectors in the cone is infinite. And you can't tell which it is.

But suppose you are Louis Guttman (1954). He says, suppose you chose one of these Y vectors as your guess. How possibly wrong could you be? He then shows you that you could be the most wrong if the correct Y is 180° directly opposite from the vector you chose. And the angle between them is twice the angle between X and your Y. Well, what would be the cosine of the angle between your guess and the correct Y? It's the cosine of twice this angle. And by a double angle formula in trignonometry we know this is equal to $\cos 2\theta = 2\cos^2\theta - 1$. So, if we convert correlations to cosines and vice versa, this means the correlation between your guess and the correct Y could be as little as $2 \times (.7071)^2 - 1 \approx 2 \times .5 - 1 = 0$. And if the correlation between X and Y is less than .7071, the correlation between your guess and the correct Y could be negative. So, your interpretation of the factor could be possibly the opposite of what it is when R is less than .7071. That gives one pause for thought.

Well, if we replace X by the regression estimate $\hat{\xi}$ of the common factor ξ, the correlation between X and Y by the multiple correlation R between the regression estimate and the unknown ξ, then the variable ξ can be any variable in the cone surrounding $\hat{\xi}$ with cosine equal to R. If the multiple correlation between $\hat{\xi}$ and ξ is $R = .95$ and the correlation between the factor estimator $\hat{\xi}$ and a variable Z is .60, what is the range of possible correlations between Z and ξ? The solution requires inverse trigonometric functions, which are available on any mathematical calculator. Let $\cos\alpha = .60$ and $\cos\rho = .95$. The minimum possible correlation is $\min r = \cos(\cos^{-1}\alpha + \cos^{-1}\rho) = \cos[\cos^{-1}(.60) + \cos^{-1}(.95)] = .3202$ and the maximum possible correlation is $\max r = \cos(\cos^{-1}\alpha - \cos^{-1}\rho) = \cos[\cos^{-1}(.60) - \cos^{-1}(.95)] = .8198$. So, our uncertainty as to what the true correlation between the latent factor and an external variable is that it is between .3202 and .8198. Furthermore, in many cases with R less than .7071 the true common factor variable could be negatively related to your interpretation of it. So, working with factor scores in common factor analysis has too much indeterminacy to be generally useful.

Factorial invariance

There are three cases of factorial invariance. (1) Invariance of the reproduced correlation matrix under rotation of factors. (2) Invariance of factors under selection of variables for analysis, and (3) invariance of model parameters under restriction of range.

Rotational invariance meant that it didn't matter what rotation for the factors took place, you would always get the same reproduced correlation matrix. Thurstone showed that there was an objective reason for picking one kind of rotation out of all, and this was the simple structure solution. Not only was it usually simpler to interpret because it disentangled the variables into clusters around certain factors, but the simple structure solution for the factors would be invariant across studies with different variables selected from the factor domain for analysis. But there is a little recognized stipulation behind that assertion. You must use care in selecting the variables to be close to the hyperplanes defining the primary axes (the factors). And this will only occur if you have a clear idea of how to define the factor domain to select from in the first place. Every variable in the factor domain is some linear combination of the same common factors of the domain, and preferably of at most $r-1$ of them. This also solves the problem of the second case, of getting the same factors with different selections of variables from the domain.

Many factor analytic studies are done with selections of subjects from restricted populations. For example, many studies use college students as subjects and these are more intelligent, more educated than many persons in the general population. It can be shown that correlations among variables become attenuated (made smaller) when computed from highly selected populations. But to make this explanation short, factor-pattern coefficients will be unaffected by restriction of range, while correlations, variances, and covariances are affected. This means factor structure coefficients are not invariant under restriction of range and difficult to meaningfully compare across studies. With comparative studies, it is important to study raw covariances rather than standardized variables with correlation coefficients within a given sample. Standardizing changes the metric of the measurments and this can become a confound with restriction of range, especially when comparing results across studies. See Mulaik (2010) for further details.

Common factor analysis versus component analysis

Common factor analysis models treat observed variables as linear functions of latent, unobserved variables. A closely related set of models are component analysis models, among which the most known is **principal components analysis**. These models find observable linear combinations of the original observed variables and treat these linear combinations as independent variables which determine the observed variables.

The model equations of a component analysis model would be

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \\ &\vdots \\ Y_1 &= a_{n1}X_1 + a_{n2}X_2 + \cdots + a_{nn}X_n \end{aligned} \tag{8.7}$$

There are no unique factor or error variables. Furthermore the independent variables $X_1, \cdots, X_n$ are themselves functions of the observed variables

$$\begin{aligned} X_1 &= b_{11}Y_1 + b_{12}Y_2 + \cdots + b_{1n}Y_n \\ X_2 &= b_{21}Y_1 + b_{22}Y_2 + \cdots + b_{2n}Y_n \\ &\vdots \\ X_n &= b_{n1}Y_1 + b_{n2}Y_2 + \cdots + b_{nn}Y_n \end{aligned} \tag{8.8}$$

Furthermore, there can be as many X variables as observed variables, but frequently there are fewer as these component analysis models seek to imitate common factor analysis models.

Mulaik (2010) describes three component analysis models: principal components (Pearson, 1901), weighted principal components (Mulaik, 1972), and image analysis (Guttman, 1953; Kaiser, 1963).

Component analysis models at first glance seem to have all the advantages and none of the disadvantages of the common factor models. They have observed independent variables constructable from the observed variables. They can have rotated factors. There is no "factor indeterminacy." Factor scores are indeed observable since they are perfectly determined by the observable variables.

One important thing that component analysis models lack is the ability to represent objective reality. Empiricists liked component analysis models because they regarded what appeared to the senses was reality. But they provided no general criterion for differentiating an illusion from reality. But science abandoned that simplistic view in favor of regarding what is real is given by a concept independent of the particular actions and means of observation of a particular observer. This concept should be coherent and invariant with diverse appearances from different points of view, to many observers, perspectives, methods of observation, from many locations, and invariant across many occasions. A dinner plate is seen as a circular, flat, and enduring object, distinct from the observer as the observer moves around and views it from different positions and even on different occasions. Some of the time its appearance is as an ellipse but it is still conceived as a round circular object.

Physicists regard certain subatomic particles as real, but no one can see them directly. One sees their effects on what is observed in a cloud chamber from different angles, as for example, the condensing of droplets of vapor in the chamber in a certain curved path typical of the particle produced by the ionizing effects on the vapor of the particle as it moves through the vapor. These particles are seen to have the same path forms in different cloud chambers, with different vapor media, different laboratories, by different researchers under similar conditions.

In common factor analysis the common and unique factors model variables exist independently of each of the observed variables in the analysis. Three or more observed variables could indicate a single common factor. Four would allow you to test for its existence, with each indicator an independent point of view of the latent variable. But Thurstone (1937, 1947) argued that unless you had a simple structure solution in a common factor analysis, you still would not have anything better than a principal components solution with a rotation to "simple structure," (like orthogonal varimax). In the latter case, the rotation would be limited to the space of just the observed variables and would differ from an analysis with communalities and unique variances. Communalities make the difference by isolating what is common to the tests and by hypothesis to other tests not in the battery. Then, using simple structure solutions, you could use different test items and still recover the same underlying abilities as common factors. Thus Thurstone held that,

> The factorial description of a test must remain invariant when the test is moved from one battery to another. This simply means that if a test calls for a certain amount of one ability and a certain amount of another ability, then that test should be so described irrespective of the other tests in the battery. (Thurstone, 1937, p. 75)

That would be a further test of the objective independence of the ability. In contrast, he said,

> If we ask the subjects to take some more tests at a later time and if we add these tests to the battery, then the principal components and centroid axes will be altered, and the factorial description of each test will also be altered. (p. 75)

The principal components maximize the variance of a linear combination of the variables. That is based on just the variables in the battery. And you cannot determine what the psychological content is of a factor without having at least three, four, or more tests or indicators of it. A single indicator may have many things within it. Having multiple indicators of it allows us to isolate what in them is common to them. And the indicators' loadings must be invariant across batteries of other tests.

But principal components and other component analysis models do produce factor-pattern matrices that are very similar in many cases to the factor-pattern matrix of a common factor analysis derived from the same data. Principal components solutions with rotations tend to overestimate the factor-pattern loadings relative to the common factor solutions. But Widaman (1993) showed that the difference in loadings between principal components analysis and common factor analysis decreased as the ratio of the number of observed variables to the number of common factors increased, when the magnitudes of the pattern loadings of the common factor model increased, and when the communalities increased. Otherwise, principal components models do not model accurately data generated by a common factor model.

Principal components models are still useful as data reduction procedures. Factor scores for a smaller number of component factors rotated to orthogonal simple structure from a principal components analysis of a given set of predictor variables are determinate and can be used to reduce the number of highly correlated predictor variables for a regression analysis and avoid effects of multicollinearity (Jolliffe, 1986).

Confirmatory Factor Analysis

Confirmatory research versus exploratory research

The philosopher of science Charles Saunders Peirce (1931–1938) was the first to expressly articulate a modern theory of scientific method that involved cycling over and over through three stages: **abduction**, **deduction**, and **induction**. In **abduction** the scientist encounters phenomena that require an explanation. He/she then considers a number of explanations that account for them and then chooses the one that best fits all of the phenomena to serve as the basis for a hypothesis. But that does not prove the validity of the hypothesis. Any number of distinct explanations might be imagined to account for the same phenomena. There is a need to test the explanation. This leads to the next stage in which **deduction** from the hypothesis together with prior knowledge and assumptions leads to an observable consequence that can be tested. The deduced consequence must not be logically possible from prior knowledge and assumptions without the hypothesis to combine with them to yield the consequence. The deduced consequence yields a specific set of quantities that should be observed if the hypothesis is true. In the third stage an appropriate set of data are collected to test the hypothesis. The data could never be all the logically possible observations one might have. And so there may be sampling error making the observed quantities deviate from the theoretical quantitities if the hypothesis were true. So, the new data should constitute a random sample from a population with a known probability distribution: then a test statistic measuring the degree of difference between the observed and hypothetical quantities should be chosen and have a known probability distribution for deviations from the hypothesized values. The hypothesis is provisionally accepted if the test statistic indicates that the deviations are not so large as to be so improbable and so deviant as to cast doubt on the truth of the hypothesis. This is an inductive inference, because the smaller and more probable the actual deviations conditional on the hypothesis, the more likely the data is consistent with the hypothesis.

If the hypothesis is accepted, other deductions leading to other new studies may be conducted to further test the hypothesis. If rejected, that starts over the abduction, deduction, and induction cycle, to discover a model that fits both previous and current data and in turn new data derived by deduction.

This represents a different way of testing substantive hypotheses in confirmatory studies. You are likely most familiar from basic statistics with nil hypothesis statistical testing. In nil hypothesis testing you begin with a substantive hypothesis that there is a difference between means or a correlation between two variables. But you have no specific value for the difference or the correlation. You just test the nil hypothesis as the null hypothesis that the difference or correlation is zero. If the result is significant, you then declare the substantive hypothesis is supported because the difference or correlation is not zero. But this is very weak support, because any number of different nonzero quantities might be hypothesized under distinct theories and be consistent with a nonzero result. In inductive inference from a deduced specific hypothesis, the null hypothesis is the prespecified theoretical quantities of a substantive hypothesis, which may be nonzero. One hopes to accept rather than reject the null hypothesis. We will now see how this applies to confirmatory factor analysis.

I need to point out here that the same kinds of procedures used by L. L. Thurstone in selecting and constructing items for his inventories to be factor analyzed also apply in confirmatory factor analysis. In many ways Thurstone tried to test theories of underlying human cognitive abilities, but all he had was the exploratory factor analysis method to do it. He would have welcomed the methods of confirmatory factor analysis developed after his death.

Model equations

Basically the model equations of confirmatory factor analysis are the same as in exploratory factor analysis as given in Equation (8.4). The difference lies in more parameters (coefficients) being constrained by hypothesis in the confirmatory factor analysis model than in the exploratory factor analysis model. Confirmatory factor analyses (Jöreskog, 1969) are done with structural equation modelling programs, because confirmatory factor analyses are special cases of more general structural equation models.

The programs for structural equations modelling (Jöreskog, 1969, 1973) indicate constraints on model parameters usually in three ways: (1) as unknown free or estimated parameters, indicated often by a *; (2) as fixed parameters having specific numerical values respectively; or (3) by constraining a group of parameters to be equal to one another. It is important to understand that free or estimated parameters are not specified and are not really what is hypothesized about the model. Not shown are the fixed zero correlations between common and unique factors and among the unique factors. These are assumed given by the model.

The free parameters are unknowns to be estimated iteratively in such a way that their values make the model fit the data as well as possible **conditional on the constraints**, which are fixed during the iterations. The estimated parameters are needed to fill out the model so that a reproduced correlation matrix among the observed variables can be generated from the model. Lack of fit will concern differences between the original, observed correlation matrix, and the reproduced correlation matrix. Any lack of fit is to be attributed to the constraints. Computer programs designed to estimate these parameters systematically adjust values for the free parameters iteratively over and over to improve fit until the reproduced model correlation matrix fits the observed correlation matrix as well as possible subject to unchanging constraints. There are no eigenvalues and eigenvectors in confirmatory factor analysis.

In the model we are testing here, we have relied on our sense of word meanings to determine what words are **not** related to what factors. We indicate this by fixed zero coefficients shown in bold type (see Table 8.7). This is usual at the outset of research.

Table 8.7 Hypothesized factor-pattern loadings are indicated by specified quantities (in this case **0**). Estimated parameters are indicated by *.

	Hypothesis						
	factor-pattern loadings						
Variable	*1*	*2*	*3*	*4*	*5*	*6*	ψ_i
ALLOW	*	**0**	**0**	**0**	*	**0**	*
OUTGO	*	*	**0**	**0**	*	**0**	*
NONTALK	**0**	*	**0**	**0**	*	**0**	*
PERMISS	*	**0**	**0**	**0**	**0**	**0**	*
FAIR	**0**	**0**	*	*	**0**	**0**	*
UNFRIEND	*	*	*	*	**0**	**0**	*
FORCELESS	*	*	**0**	*	*	**0**	*
RESTRAIN	*	*	**0**	**0**	*	**0**	*
NONDOMIN	*	*	**0**	*	*	**0**	*
UNLOVING	**0**	**0**	**0**	*	**0**	**0**	*
NONCNFRM	**0**	**0**	**0**	**0**	**0**	*	*
COMPLINT	*	**0**	**0**	**0**	**0**	*	*
IMPARTIAL	**0**	*	*	**0**	*	**0**	*
LEADER	**0**	**0**	**0**	**0**	*	**0**	*
UNKIND	**0**	*	*	*	**0**	**0**	*
UNCNVENT	**0**	**0**	**0**	**0**	**0**	*	*
DIRECTNG	**0**	**0**	**0**	*	*	**0**	*
PROHIBTV	*	**0**	**0**	**0**	*	*	*
UNGREGAR	**0**	*	**0**	**0**	**0**	**0**	*
JUST	**0**	**0**	*	*	**0**	**0**	*
NEUTRAL	*	*	*	*	*	**0**	*
SOCIABLE	*	*	**0**	*	*	*	*
UNAFFECT	**0**	*	**0**	*	**0**	**0**	*
REBELLOS	**0**	*	**0**	*	**0**	*	*

	Correlations among factors					
	1	*2*	*3*	*4*	*5*	*6*
Permissive 1	**1.00**					
Introversion 2	*	**1.00**				
Fair, Just 3	*	*	**1.00**			
Unloving 4	*	*	*	**1.00**		
Nondirecting 5	*	*	*	*	**1.00**	
Unconventional 6	*	*	*	*	*	**1.00**

We do not know what the loadings are of certain variables on certain factors, so we leave them as estimated free parameters. We may believe they have a nonzero loading, but by not knowing what these specifically are, we are compelled to leave them as estimated unknowns. If a factor is to be determined by the hypothesized model it must be in terms of what variables are not related to the factor. Thus the hypothesis tested is about the particular pattern of zero coefficients in the factor loading matrix in the context of the fixed zero correlations between common and unique factors and among the unique factors.

In later research, after we have found specific nonzero loadings for certain variables on certain factors, we may in addition to zero loadings, specify these nonzero values as fixed parameters, while leaving the common factor variances to be free parameters, which will fix the metric for the factors as the same as in an initial study. In such later studies we should also use raw covariances instead of correlations for the observed variables. "Fixing the metric" means specifying the units of measurement in terms of how much a unit change of one variable yields a given change in another variable. One parameter needs to be fixed for each common factor, either its variance or the loading of a selected indicator on it to determine the metric. Because in this study both common factors and observed variables are standardized to have unit variances, we would have to rely on the fixed value of a factor loading to establish the metric in later studies while letting observed variables and common factors have whatever variances the data allows.

In Table 8.8, I show the results of a confirmatory factor analysis. We show the common factor-pattern loadings and the unique factor loading for each variable, indicated by name on the left. The correlation matrix among the six factors is also shown. The largest nonzero loadings in each column correspond to items thought by hypothesis to best represent a corresponding factor. While that may be thought of as a test of the model, the test is really about the fixed zero coefficients, and that includes the zero coefficients representing zero correlations between common and unique factor variables, and among the unique factor variables as well.

Tests of fit

What we need to consider now is an overall test of the fit of the model to the data, which is the chi-square test. The chi-square value was 406.672 with 201 degrees of freedom. For a sample size of $N = 523$, the p value is less than .001, which indicates a statistical lack of fit. The sample size allows for high power to detect small discrepancies. But the model still has a close approximation to the data in the fit of the reproduced model correlation matrix to the observed correlation matrix, indicated by a CFI index of .958, where 1.00 would indicate perfect fit. The close approximation means just that. It does not mean the model is necessarily exactly correct. The significant chi-square indicates a need to discover possible reasons for the failure to get fit to within sampling error. There are numerous possibilities to consider. To begin with, the correlations of the observed variables were assumed to be among continuous variables distributed as the multivariate normal distribution. But we know they were **discrete** polytomous variables. That alone could produce some lack of fit in the discrepancy between distributional assumptions and reality.

Table 8.8 Results of confirmatory factor analysis of the 24 trait words based on six hypothesized concepts.

	Results						
	factor-pattern loadings						
Variable	*1*	*2*	*3*	*4*	*5*	*6*	ψ_i
ALLOW	.805	**0**	**0**	**0**	−.049†	**0**	.570
OUTGO	.225	−.660	**0**	**0**	.183	**0**	.183
NONTALK	**0**	.814	**0**	**0**	.081†	**0**	.631
PERMISS	.724	**0**	**0**	**0**	**0**	**0**	.689
FAIR	**0**	**0**	.756	−.064	**0**	**0**	.601
UNFRIEND	−.163	.271	−.165	.395	**0**	**0**	.665
FORCELESS	.353	.276	**0**	−.304	−.399	**0**	.589
RESTRAIN	−.071†	.412	**0**	**0**	.355	**0**	.914
NONDOMIN	.231	.264	**0**	−.381	−.495	**0**	.587
UNLOVING	**0**	**0**	**0**	.804	**0**	**0**	.595
NONCNFRM	**0**	**0**	**0**	**0**	**0**	.495	.869
COMPLINT	.174	**0**	**0**	**0**	**0**	−.508	.839
IMPARTIAL	**0**	.167	.362	**0**	−.095	**0**	.925
LEADER	**0**	**0**	**0**	**0**	.799	**0**	.601
UNKIND	**0**	.031†	−.216	.662	**0**	**0**	.556
UNCNVENT	**0**	**0**	**0**	**0**	**0**	.707	.707
DIRECTNG	**0**	**0**	**0**	.261	.834	**0**	.573
PROHIBTV	−.491	**0**	**0**	**0**	−.001†	−.103	.868
UNGREGAR	**0**	.357	**0**	**0**	**0**	**0**	.934
JUST	**0**	**0**	.836	−.033†	**0**	**0**	.514
NEUTRAL	.144	.212	.317	.183	−.044†	**0**	.844
SOCIABLE	.241	−.465	**0**	−.135	.193	−.101	.696
UNAFFECT	**0**	.060†	**0**	.721	**0**	**0**	.665
REBELLOS	**0**	−.206	**0**	.301	**0**	.575	.696

	Correlations among factors					
	1	*2*	*3*	*4*	*5*	*6*
Permissive 1	**1.00**					
Introversion 2	†	**1.00**				
Fair, Just 3	−.377	−.176	**1.00**			
Unloving 4	−.535	.392	−.650	**1.00**		
Nondirecting 5	−.319	−.525	.291	−.213	**1.00**	
Unconventional 6	−.046†	−.015†	−.398	.282	−.138	**1.00**

Note. **Bold type** indicates parameters prespecified and fixed by hypothesis. Numbers not in bold type are estimated values. † indicates an estimated value that does not differ significantly from zero. Zero correlations between common and unique factors and among unique factors are not shown.

But substantively each one of the fixed zero coefficients in the hypothesized model could be responsible for lack of fit. There are individual single degree-of-freedom tests for each of these fixed parameters separately known as Lagrange multiplier tests. Associated with each is a p value. These tests are not statisticially independent, so after the first test of the largest chi-square value, the p values may be off the mark. I will not show those here, other than to mention that the largest significant discrepancy with a chi-square of 4.889 and 1 degree of freedom was on a correlation between the unique factors of the FORCELESS and the NONDOMINATING variables. Both share (negatively) the common factor of DIRECTING with LEADER, DIRECTING, and RESTRAINING. All of these concern some aspect of applying force in the interpersonal realm. But people who are forceless and nondominating may have other characteristics in common which would come out with these variables on what is now a possible doublet factor. But that is only the basis for a hypothesis that might be tested in the future by including other variables such as indicators of anxiety, timidity, inadequacy, physical strength, and size to see if these also contribute to the common variance of these two variables.

In any case, a careful examination of the results is essential to discover possible explanations of the discrepancy needing further testing in future research.

Tests are also possible for the free parameters. The standard errors are usually given for these in structural equation modelling programs used to do confirmatory factor analyses. So, it is possible to test prior nonzero hypotheses as to their values. But ordinarily, you might simply test the hypothesis that these values differ significantly from zero. In that case the programs usually provide the z-statistic based on the standardized normal distribution for the test of the difference from zero. A z value greater than 2.00 would indicate significance. In Table 8.8 I have used the symbol † to show that a loading treated as a free parameter, was estimated to be not significantly different from zero.

It is also important to learn that degrees of freedom respresent the degree to which the model is disconfirmable by lack of fit. Mulaik (2001, 2009) showed that the number of degrees of freedom equaled the number of dimensions in which the model of the data was free to differ from the data and thus fail to fit. It is given in factor analysis and structural equation modelling by the formula

$$df = n(n+1)/2 - m \tag{8.9}$$

where df are the degrees of freedom of the model, $n(n+1)/2$ is the number of distinct elements of the observed symmetric correlation/covariance matrix to be fitted, and m the number of estimated parameters. Thus for the present data we have $24(24+1)/2-[24+6(5)/2+60]=201$. There are 300 elements to fit to. There are 24 unique variances, 15 correlations among 6 factors, and 60 factor-pattern coefficients to estimate. Hence there are 201 degrees of freedom. The more degrees of freedom, the more disconfirmable is the model. Given the number of elements to fit, you lose a degree of freedom for each parameter estimated. Hence ideally you should estimate as few parameters as possible. And models with more estimated parameters are able to better fit the data just as a mathematical artifact. So, estimating lots of parameters is not desirable. A scientific model is more desirable the more ways by which it can be disconfirmed by data. If it passes the fit test, the more degrees of freedom indicate the stronger the evidence for the model. An exploratory factor analysis model usually has fewer degrees of freedom than a corresponding confirmatory factor analysis model because it estimates more parameters.

Some Remaining Problems.

Factorial homogeneity

The common factor model, being the oldest multivariate technique used by psychologists, has achieved widespread use. But an almost universally unrecognized assumption of its use is that every subject should behave according to the same common factor model in responding to tests or questionaires. I call this the implicit assumption of **factorial homogeneity of the observations**. This assumes that one subject or observation can be substituted for another without changing the factor analytic results. This does not assume that they will get scores at the same level on each variable, but that their scores will be the same linear functions of the same common factors, regardless of the level of their scores on the factors.

How might this assumption be compromised? In the case of personality trait ratings and attitude questionnaires, subjects might vary in their understanding and use of the words. As a result the combined ratings of all raters in computing the variances and correlations may yield factor loadings and correlations that represent no single individual rater exactly. The model may then fail to fit precisely and yield significant chi-square goodness of fit indices. On performance tests, subjects may vary in the manner in which they perform a task or solve a problem. Those at different levels of the spectrum of ability may use different abilities to try to solve the same problem implying different factor analytic models are needed at different levels of ability. Unlike physical objects, humans are highly variable and complex in their behavior.

Factorial heterogeneity

The opposite case I call **factorial heterogeneity of the observations**. How can we determine whether this is the case with a given sample of responders to our rating scales and performance tests? Can we isolate individuals in a sample into subgroups that are factorially homogeneous? That I leave as a problem for some doctoral dissertation in the future to resolve. But it is an important problem yet unresolved.

Nonlinearity

Another problem concerns the possible nonlinearity of functional relations between variables. Factor analysis assumes linear functional relations between factors and observed variables. But suppose, in fact, the relations between factors and observed variables are monotonically increasing (or decreasing). A monotonically increasing (decreasing) function is one in which, as the independent variable increases (decreases), the dependent variable never decreases (increases) but always increases (decreases). But the increase (decrease) may not be by constant sized increments. Monotonic functions can be approximated in many cases with straight line (linear) functional relations. But at some point there will be lack of fit. More serious is the case in which nonlinear nonmonotonic functional relations exist between observed dependent variables and the independent common factor variables wherein the dependent variables are increasing and decreasing throughout the range of the independent variables. Factor analysis with linear relations cannot model this behavior, even approximately. Factor analysis fitted to such data may require many more common factors than actually are driving the behavior of the observed variables.

Causal structure

Another problem is that factor analysis only assumes that any relations between two observed variables are due to a third variable common to them: $\Upsilon_1 \leftarrow \xi_j \rightarrow \Upsilon_2$. But other possibilities are that one variable is the cause of the other, $\Upsilon_1 \rightarrow \Upsilon_2$ or $\Upsilon_1 \leftarrow \Upsilon_2$. Two observed variables may be correlated because there is an unobserved chain of latent variables linking them, $\Upsilon_1 \rightarrow \xi_j \rightarrow \xi_j \rightarrow \Upsilon_2$. Factor analysis is not able to reveal such relationships. Since the 1970s new models with latent and observed variables have developed to handle such relationships, known as structural equations models. But there are many textbooks now available on that subject.

In the twenty-first century, factor analysis will still have a place in some studies where the data conforms to a common factor model. But the rapid development of linear and nonlinear models of complex causal relations between variables will mean these will be increasingly the focus of researchers in this century.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*, 676–696.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–115.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge: Cambridge University Press.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurment, 15*, 139–164. DOI: 10.1111/j.1745-3984.1978.tb00065.x.

Galton, F. (1886). Family likeness in stature. *Proceedings of the Royal Society, 40*, 42–73.

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings Royal Society of London, 45*, 135–145.

Galton, F. (1889). *Natural Inheritance.* London: Macmillan.

Guilford, J. P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika, 18*, 277–296.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149–161.

Guttman, L. (1956). "Best Possible" systematic estimates of communalities. *Psychometrika, 21*, 273–285.

Guttman, L. (1965). A faceted definition of intelligence. In R. R. Eiferman (Ed.), *Studies in psychology, Scripta hierosolymitana* (Vol. *14*). Jerusalem, Israel: The Hebrew University.

Guttman, L. (1994). *Louis Guttman on theory and methodology: Selected writings.* S. Levy (Ed.). Aldershot, UK: Darthmouth Publishing Co.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.
Horn, J. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattel (Eds.), *Handbook of multivariate experimental psychology* (2nd ed.) (pp. 645–665). New York: Plenum Press.
Howe, W. G. (1955). Some contributions to factor analysis. Report No. ONRL-1919. Oak Ridge, TN: Oak Ridge National Laboratory.
Humphreys, L. G., Tucker, L. R., & Dachler. P. (1970). Evaluating the importance of factors in any given order of factoring. *Multivariate Behavioral Research, 5*, 209–215.
Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika, 67*, 7–20.
Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika, 71*, 173–191.
Jennrich, R. I., & Sampson, P. F. (1966). Rotation to simple loadings. *Psychometrika, 31*, 313–323.
Jolliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.
Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443–482.
Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183–202.
Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 86–223). New York: Seminar Press.
Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200.
Kaiser, H. F. (1961). A note on Guttman's lower bound for the number of common factors. *British Journal of Statistical Psychology, 14*, 1–2.
Kaiser, H. F. (1963). Image Analysis. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 156–166). Madison, WI: University of Wisconsin Press.
Katz, J. O., & Rohlf, F. J. (1974). Functionplane – A new approach to simple strucure rotation. *Psychometrika, 39*, 37–51.
Kiers, H. A. L. (1994). SIMPLIMAX: Oblique rotation to an optimal target with simple structure. *Psychometrika, 59*, 567–579.
Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research, 34*, 347–365.
McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin, 107*, 247–255.
McKeon, J. J. (1968). *Rotation for maximum association between factors and tests* (Unpublished manuscript). Biometric Laboratory, George Washington University.
Mulaik, S. A. (1963). *A factor analytic investigation of the equivalence of personality factors with semantic factors* (Unpublished doctoral dissertation). University of Utah, Salt Lake City, UT.
Mulaik, S. A. (1964). Are personality factors raters' conceptual factors? *Journal of Consulting Psychology, 28*, 506–511.
Mulaik, S. A. (1996a). On Maraun's deconstructing of factor indeterminacy with constructed factors. *Multivariate Behavioral Research, 31*, 579–592.
Mulaik, S. A. (1996b). Factor analysis is not just a model in pure-mathematics. *Multivariate Behavioral Research, 31*, 655–661.
Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw Hill.
Mulaik, S. A. (2001). The curve-fitting problem: An objectivist view. *Philosophy of Science, 68*, 218–241.
Mulaik, S. A. (2005). Looking back on the factor indeterminacy controversies in factor analysis. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 173–206). Mahwah, NJ: Lawrence Erlbaum Associates.

Mulaik, S. A. (2009). *Linear causal modeling with structural equations.* Boca Raton, FL: Chapman and Hall/CRC Press, Taylor & Francis Group.
Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC Press, Taylor & Francis Group.
Mulaik, S. A., & McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika, 43*, 177–192.
Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Champaign, IL: University of Illinois Press.
Pearson, K. (1896). Mathematical contributions to the theory of evolution, III: regression, heredity and panmixia. *Philosophical Transactions of the Royal Society of London (A), 18*, 253–318.
Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*, 559–572.
Peirce, C. S. (1931–1958). *Collected papers of Charles Sanders Peirce.* (Vol. 1–8); C. Hartshorne & P. Weiss (Eds.) (Vol. 1–6), and A. W. Burks (Ed.) (Vols. 7–8). Cambridge, MA: Harvard University Press.
Rozeboom, W. W. (1991a). Hyball: A method for subspace constrained factor rotation. *Multivariate Behavioral Research, 26*, 163–177.
Rozeboom, W. W. (1991b). Theory and practice of hyperplane optimization. *Multivariate Behavioral Research, 26*, 179–197.
Rushton, J. P., & Irwing, P. (2011). The general factor of personality: normal and abnormal. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences.* New York: Wiley-Blackwell.
Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53–61.
Spearman, C. (1904a). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.
Spearman C. (1904b). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201–293.
Spearman, C. (1927). *The abilities of man.* New York: Macmillan.
Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, IO.
Thurstone, L. L. (1937). Current misuse of the factorial methods. *Psychometrika, 2*, 73–76.
Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago, IL: University of Chicago Press.
Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement, 58*, 541–568.
Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.
Webb, E. (1915). *Character and intelligence: an attempt at an exact study of character.* Cambridge: Cambridge University Press.
Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*, 263–311.
Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis.* Albany: State University of New York Press.
Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society, 60*, 812–854.
Yule, G. U. (1900). On the association of attributes in statistics: with illustrations from the material of the Childhood Society, &c. *Philosophical Transactions of the Royal Society of London (A), 194*, 257–319.