

PRINCIPAL-AXIS FACTOR ANALYSIS

Authored by
mohammad looti

October 25, 2025

RECOMMENDED CITATION

mohammad looti (2025). *PRINCIPAL-AXIS FACTOR ANALYSIS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=55032>

PRINCIPAL-AXIS FACTOR ANALYSIS

Primary Disciplinary Field(s): Statistics, Psychometrics, Data Science

1. Core Definition

Principal-Axis Factor Analysis (PAFA), often referred to interchangeably with **Common Factor Analysis**, is a statistical method employed primarily within the behavioral and social sciences to uncover the underlying structure or latent factors responsible for the correlations observed among a large set of measured variables. Unlike related dimensional reduction techniques, such as Principal Component Analysis (PCA), PAFA operates under the fundamental assumption that the variance observed in any given variable can be decomposed into two distinct parts: the variance shared with other variables in the set (known as **common variance** or communality), and the variance unique to that specific variable (known as unique variance, encompassing both measurement error and specificity). PAFA's essential goal is to determine the smallest number of hypothetical, unobserved factors that can adequately account for this shared variance, thereby simplifying the complex relationships observed in the data structure.

The mathematical objective of PAFA is to model only the common variance, seeking to remove or disregard the unique variance inherent in each observed measure. This is rooted in the belief that only the shared variance reflects meaningful, underlying psychological or physical constructs. Therefore, PAFA attempts to estimate a parsimonious set of factors--the principal axes--that explain the covariance structure among the variables. The result is a factor structure where variables that load highly onto the same factor are presumed to be measuring the same latent construct. This process is crucial for the development and validation of measurement scales, theoretical model building, and data compression when focusing strictly on the causal mechanisms reflected by the common traits.

2. Etymology and Historical Development

The origins of factor analysis date back to the early 20th century, largely spurred by the work of British psychologist Charles Spearman (1904), who developed the two-factor theory of intelligence. Spearman posited that intelligence could be largely explained by a single general factor (the 'g' factor) and specific factors ('s' factors) unique to individual tests. This early model explicitly separated common variance (g) from unique variance (s), laying the theoretical groundwork for the PAFA approach. Spearman's model emphasized the necessity of analyzing only the shared components of variability among measures, distinguishing it fundamentally from subsequent methods that accounted for total variance.

Throughout the subsequent decades, researchers like L.L. Thurstone refined and expanded these

concepts, moving beyond the single-factor model to develop methods for extracting multiple common factors, culminating in his work on Multiple Factor Analysis in the 1930s. The methodological difficulty in performing Common Factor Analysis, however, lay in the challenge of precisely estimating the communalities--the amount of variance each variable shares with the latent factors. Unlike PCA, which assumes communalities of 1.0 (treating all variance as common), PAFA requires an iterative estimation process. Prior to the advent of modern computers, the intensive calculation required for iterative solutions often made simpler methods, such as PCA, more practical for researchers.

It was not until the mid-to-late 20th century that the computational power became readily available to implement the complex, iterative algorithms necessary for PAFA effectively. This accessibility allowed PAFA to become a standard tool in psychometric research, particularly when the goal was theoretical validation and the construction of instruments designed to measure true, stable latent traits, rather than simply reducing the dimensionality of a dataset for convenience. The ability of PAFA to explicitly model and isolate measurement error is its enduring methodological advantage, ensuring that the resulting factors are theoretically purified estimates of the underlying constructs.

3. Key Characteristics

A defining characteristic of Principal-Axis Factor Analysis is its focus on the structure of the correlation matrix after initial variances have been adjusted for uniqueness. This process requires the substitution of initial estimates for the communalities (h^2) along the main diagonal of the correlation matrix, replacing the perfect unity (1.0) typical of PCA. These initial communality estimates are crucial, as they determine the initial factor space. A common method for initial estimation is using the squared multiple correlation (SMC) of each variable with all other variables in the dataset. This technique provides a lower-bound estimate of the true communality, reflecting the shared variance derived from the linear relationships among the measured items.

The resulting factor model is structured around the distinction between common and unique factors. The mathematical representation of an observed variable (Z_i) in PAFA is:

$$Z_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + U_i$$

Where F_1 through F_m represent the **common factors** (the latent traits being sought), a_{ij} are the factor loadings (the correlation between the variable and the factor), and U_i represents the **unique factor**. This unique factor (U_i) captures all variability not shared with the common factors, encompassing both the reliable variance specific to that measure (specificity) and random measurement error. The emphasis on U_i is the cornerstone of PAFA's design, distinguishing it as a true factor analytic model aiming to explain variance caused by latent constructs.

Furthermore, PAFA is inherently an **iterative method**. After the initial factors are extracted based

on the starting communality estimates, the algorithm recalculates the communalities based on the extracted factors (the sum of the squared loadings for each variable). These new estimates are then fed back into the correlation matrix, and the factors are re-extracted. This process continues until the communality estimates stabilize and the difference between successive iterations falls below a predetermined convergence criterion. This iterative refinement is essential for achieving the most accurate and theoretically sound representation of the common factor space, differentiating PAFA from non-iterative techniques like PCA, which provide a single, fixed solution based on total variance.

4. Methodology and Steps

The implementation of Principal-Axis Factor Analysis involves a rigorous sequence of statistical steps designed to isolate and define the common factor structure. The process begins with data preparation and the calculation of the correlation matrix, followed by the specific steps of extraction, determination, and interpretation.

A. Communality Estimation

The first critical methodological step unique to PAFA is the initial estimation of **communalities**. As noted, the main diagonal of the correlation matrix is replaced with estimates of the proportion of variance in each variable explained by the common factors. Common techniques include using the Squared Multiple Correlation (SMC), which is the R^2 obtained by regressing the variable against all other variables in the set. This initial estimate is crucial as it determines the initial size of the common factor space available for extraction.

B. Factor Extraction

Once the matrix is adjusted, the factors are extracted using mathematical techniques, often involving the eigenvalues and eigenvectors of the reduced correlation matrix. The goal of this phase is to extract factors sequentially, maximizing the variance explained by the first factor, then the second factor from the remaining variance, and so on. The primary decision at this stage is determining the optimal number of factors to retain. Standard criteria for retention include the **Kaiser criterion** (retaining factors with eigenvalues greater than 1.0) and the **scree plot test**, which visually identifies the point where the eigenvalues level off.

C. Factor Rotation

Following extraction, the initial factor solution is often mathematically correct but difficult to interpret, as most variables may load moderately onto multiple factors. Therefore, **factor rotation** is applied to achieve a simpler, more interpretable structure, following Thurstone's principle of Simple Structure. Rotation aims to maximize high loadings on some factors while minimizing

loadings on others, making the factors distinct and meaningful. There are two primary types of rotation: **Orthogonal rotation** (e.g., Varimax), which assumes the factors are uncorrelated, and **Oblique rotation** (e.g., Promax, Oblimin), which allows the factors themselves to be correlated. Oblique rotation is generally preferred when analyzing psychological constructs, as latent traits are rarely perfectly independent.

5. Significance and Impact

Principal-Axis Factor Analysis holds immense significance across various quantitative disciplines, particularly in psychometrics, sociology, and marketing research, due to its ability to test and refine theoretical models of latent structure. Its primary impact lies in its superior capacity for **construct validation**. When researchers develop a new psychological scale (e.g., measuring anxiety, job satisfaction, or personality traits), PAFA provides the rigorous framework necessary to confirm whether the items designed to measure a specific construct actually group together empirically, thus confirming the internal validity of the scale.

Furthermore, PAFA is pivotal in **theory generation and refinement**. By revealing a parsimonious set of underlying factors, PAFA can transform a large, complex dataset of observed variables into a compact, meaningful theoretical model. For instance, in personality psychology, PAFA was instrumental in reducing hundreds of trait descriptors into the robust and widely accepted Five-Factor Model (the Big Five), confirming that five primary latent traits account for most personality variance. This ability to move beyond mere data reduction and directly test the hypothesized structure of underlying constructs makes PAFA a foundational tool for advancing theoretical knowledge.

In practical terms, PAFA enables **data interpretation and reporting** that is cleaner and more theory-driven. By focusing exclusively on common variance, the factors extracted are conceptually purer estimates of the latent constructs, uncontaminated by measurement noise specific to individual items. This robustness makes the resulting factor scores, often used as composite variables in subsequent analyses (such as regression or ANOVA), better representations of the true latent variables being studied, thereby increasing the reliability and generalizability of research findings.

6. Debates and Criticisms

Despite its widespread acceptance, Principal-Axis Factor Analysis is subject to several methodological and philosophical debates. A major point of contention revolves around the inherent **subjectivity** of several key decisions that must be made by the researcher. These include the choice of the communality estimation method, the criterion used to determine the optimal number of factors to retain (e.g., Kaiser criterion vs. scree test, which often yield different results),

and the selection of the factor rotation technique (orthogonal vs. oblique). These choices are not always statistically dictated and can significantly alter the resulting factor structure and subsequent interpretation.

A significant criticism often leveled at PAFA, and factor analysis generally, concerns the **sample size requirements**. Factor analytic techniques require large samples to ensure stable correlation estimates and reliable factor structures. Rules of thumb often suggest a minimum of 5-10 participants per variable, or an absolute minimum sample size of 100 or 200, depending on the complexity of the structure and the magnitude of the expected loadings. Insufficient sample size can lead to unstable and non-replicable factor solutions, undermining the validity of the structural findings.

Finally, there is an ongoing philosophical debate regarding the **reality of latent factors**. Critics argue that while PAFA provides a mathematically efficient way to summarize correlations, the factors are merely mathematical constructions and may not correspond to genuine, observable psychological or physical entities. The interpretation of a factor is heavily reliant on the researcher's theoretical lens and judgment, meaning that different researchers analyzing the same dataset might label and interpret the extracted factors differently, leading to ambiguity in the theoretical conclusions drawn from the analysis.

7. Further Reading

[Factor Analysis \(Wikipedia\)](#)

[StatSoft: Principal Components vs. Factor Analysis](#)

[Gorsuch, R. L. \(1983\). Factor Analysis \(2nd ed.\). Lawrence Erlbaum Associates.](#)

[Brown, T. A. \(2015\). Confirmatory Factor Analysis for Applied Research \(2nd ed.\). Guilford Press.](#)