

# FACTOR ANALYSIS

Authored by  
**mohammad looti**

October 15, 2025

## RECOMMENDED CITATION

mohammad looti (2025). *FACTOR ANALYSIS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=48116>

## FACTOR ANALYSIS

**Primary Disciplinary Field(s):** Statistics, Psychometrics, Multivariate Data Analysis, Social Sciences

### 1. Core Definition and Purpose

Factor Analysis (FA) is a multivariate statistical technique primarily employed for **dimensionality reduction** and the exploration of underlying latent structures within a complex dataset. The fundamental objective of FA is to take a large, observed set of variables (e.g., test scores, questionnaire responses, or economic indicators) and distill them down into a much smaller, manageable set of inferred, unobserved constructs, known as **factors** or latent variables. This process is crucial in research design, as highlighted in the provided source material, allowing researchers to move from measuring hundreds of specific items to quantifying a few key, theoretical concepts that are responsible for the observed covariation among those items.

The mathematical foundation of Factor Analysis posits that the observed correlation among measured variables is due entirely to their shared relationship with these underlying factors. For instance, if a large battery of psychological tests measuring different aspects of cognitive ability shows high inter-correlation, FA suggests that a single, latent factor--perhaps **General Intelligence (g)**--is driving these scores. By reducing redundancy and focusing on these latent dimensions, Factor Analysis aids in parsimony, helping to create more efficient models and refined theoretical frameworks. It transforms a broad, often unwieldy collection of specific measurements into a succinct set of powerful explanatory variables.

Fundamentally, FA operates by analyzing the structure of correlations among variables. It determines the minimum number of hypothetical constructs necessary to account for the maximum amount of common variance observed in the measured variables. This procedure effectively filters out the 'noise'--the variance unique to each specific variable--and focuses only on the **common variance** that reflects the influence of the shared, latent factors. The output of this procedure is a set of factor loadings, which are coefficients indicating the strength and direction of the relationship between the observed variables and the extracted factors, thus providing the empirical basis for interpreting the meaning of the latent constructs.

### 2. Etymology and Historical Development

The conceptual origins of Factor Analysis are deeply rooted in the field of Psychometrics, particularly in the study of human intelligence. The technique is often credited to the British psychologist Charles Spearman, who, in the early 20th century (specifically 1904), developed the framework for two-factor theory. Spearman sought to demonstrate that scores across various

mental tests were all influenced by a single, common latent variable he termed the **general factor of intelligence (g)**, alongside specific factors (s) unique to each test. Spearman's early methodology provided the necessary algebraic tools for quantifying these relationships and isolating the variance attributable to g, establishing the precedent for all subsequent factor analytic methods.

The methodology was significantly expanded and formalized by American psychologist Louis L. Thurstone in the 1930s. While Spearman focused exclusively on extracting a single general factor, Thurstone introduced the concept of **multiple factor analysis**. Thurstone argued that intelligence, and psychological traits more broadly, were composed of several distinct, primary mental abilities rather than a singular general factor. His work introduced critical methodological advancements, including the concept of factor rotation, which aimed to achieve the criterion of **simple structure**--making the interpretation of factors cleaner by maximizing high loadings on one factor and minimizing loadings on others. Thurstone's contributions transitioned FA from a specialized tool for intelligence research into a general-purpose multivariate method.

The practical application and computational feasibility of Factor Analysis exploded with the advent of high-speed digital computers in the mid-20th century. Until that time, the complex matrix algebra required for solving factor solutions made large-scale FA prohibitively difficult. The development of sophisticated algorithms, coupled with increased computational power, allowed researchers to explore complex structures involving hundreds of variables and large sample sizes. Furthermore, the 1970s and 1980s saw the critical distinction solidified between **Exploratory Factor Analysis (EFA)**, used for discovering structures, and **Confirmatory Factor Analysis (CFA)**, used for testing pre-specified theoretical structures, integrating FA seamlessly into the broader framework of Structural Equation Modeling (SEM).

### 3. Mathematical Foundations: The Factor Model

At its core, Factor Analysis relies on the fundamental algebraic factor model. This model expresses each observed variable as a linear combination of the underlying common factors and a unique factor (error term). Mathematically, the model for an observed score ( $X_i$ ) is represented as:  $X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{ik}F_k + U_i$ , where  $F_1$  through  $F_k$  are the unobserved common factors,  $a_{i1}$  through  $a_{ik}$  are the **factor loadings** (the regression weights of the variable on the factor), and  $U_i$  is the unique factor variance. The goal of the analysis is to estimate these loadings and the variance associated with the factors.

A crucial distinction in the mathematical model is the partitioning of variance. The total variance of any observed variable ( $X_i$ ) is decomposed into two major parts: **Communality** and **Uniqueness**. Communality ( $h_i^2$ ) represents the proportion of variance in the observed variable that is accounted for by the common factors extracted in the analysis. This is the shared

variance that ties the variable conceptually to the latent construct. Conversely, Uniqueness ( $u_i^2$ ) is the portion of variance not explained by the common factors, encompassing both specific variance (variance unique to that measure, not shared by others) and measurement error. Factor Analysis explicitly seeks to maximize communality, thus focusing the analysis strictly on the interrelationships driven by the underlying constructs.

The estimation process typically begins with the analysis of the **correlation matrix** of the observed variables. The core task of the FA algorithm (whether it be Principal Axis Factoring, Maximum Likelihood, or others) is to estimate the communalities and produce the factor loading matrix ( $A$ ) that, when multiplied by its transpose ( $A A^T$ ), best reproduces the original observed correlation matrix ( $R$ ). The discrepancy between the observed correlation matrix and the correlation matrix implied by the factor model is minimized through iterative procedures, ensuring the resulting factors provide the most parsimonious and best-fitting representation of the data structure. The eigenvalues associated with the factors are key statistics, representing the amount of variance explained by each factor, guiding the decision on how many factors to retain.

#### 4. Key Analytical Methods and Extraction Techniques

While the term Factor Analysis is often used generically, it encompasses several distinct extraction techniques, each relying on different statistical assumptions and objectives. The most common extraction method is **Principal Components Analysis (PCA)**, although strictly speaking, PCA is a data reduction technique rather than a true Factor Analysis. PCA aims to account for the maximum amount of total variance (including both common and unique variance) in the dataset by creating orthogonal linear combinations of the variables. Researchers often confuse PCA with Factor Analysis because they yield similar results under certain conditions, but their theoretical goals differ significantly: PCA seeks optimal data summarization, while true Factor Analysis (e.g., Principal Axis Factoring) seeks to model the common variance due to latent constructs.

A true Factor Analysis technique, such as **Principal Axis Factoring (PAF)**, is designed specifically to estimate the common factors. Unlike PCA, PAF requires prior estimates of the communalities (i.e., the proportion of variance shared among variables) before factor extraction can begin. PAF iteratively refines these communality estimates until the resulting factor structure stabilizes, making it a powerful tool for theory testing where the identification of underlying latent variables is paramount. Other extraction methods include **Maximum Likelihood (ML) Factor Analysis**, which is preferred when variables are assumed to follow a multivariate normal distribution. ML provides a statistical test for assessing the goodness-of-fit of the hypothesized number of factors, offering a rigorous inferential basis often lacking in descriptive methods like PCA or PAF.

The choice of extraction method is critical and depends heavily on the researcher's goals and the nature of the data. For instance, if the primary goal is simply to compress the data set into fewer

variables for use in subsequent regression analysis, PCA is often sufficient and computationally simpler. However, if the goal is to develop and validate a psychological scale--where the focus must be on isolating the theoretical constructs responsible for the observed covariances--then PAF or ML factoring is statistically more appropriate. Regardless of the method chosen, the critical decision of how many factors to retain relies on various criteria, including the **Kaiser criterion** (retaining factors with eigenvalues greater than 1), visual inspection using a **scree plot**, or parallel analysis, all designed to determine the point at which subsequent factors explain only trivial amounts of residual variance.

## 5. The Process of Factor Rotation and Interpretation

Once the factors have been extracted, the initial factor solution often produces factors that are mathematically optimal but difficult to interpret meaningfully in a theoretical context. This necessity leads to the stage of **factor rotation**, a procedure that geometrically transforms the factor axes to achieve a **simple structure**, as originally proposed by Thurstone. Rotation does not change the mathematical fit of the model to the data or alter the communalities; it merely reallocates the variance explained among the factors, making the pattern of loadings easier to interpret. The goal is for each variable to load highly on only one factor and near zero on all others, thereby clearly demarcating the content and boundaries of each latent construct.

Factor rotation methods are broadly categorized into two types: **Orthogonal Rotation** and **Oblique Rotation**. Orthogonal methods (such as Varimax, Quartimax, and Equamax) maintain the assumption that the extracted factors are completely uncorrelated with one another--that is, the factor axes remain at 90-degree angles. **Varimax** is the most widely used orthogonal rotation, as it focuses on simplifying the columns of the loading matrix (the factors) by maximizing the variance of the squared loadings within each factor, leading to a cleaner delineation of independent constructs. Orthogonal rotation is often preferred for its simplicity and when theoretical expectation dictates that the underlying traits should be independent.

In contrast, **Oblique Rotation** (such as Promax or Oblimin) allows the resulting factors to be correlated. This is often more theoretically realistic, especially in social and behavioral sciences, where constructs like anxiety, depression, and stress are expected to be related rather than entirely independent. Oblique rotation produces two matrices: the factor loading matrix (Pattern Matrix) and the factor correlation matrix. The factor correlation matrix provides essential insight into the relationships among the latent constructs themselves, offering a richer theoretical description. The interpretation process requires careful consideration of the pattern matrix, where the researcher assigns theoretical meaning to the factor based on the variables that load highest upon it, often resulting in the naming of the newly defined latent variables.

## 6. Types and Applications in Research

Factor Analysis is applied across numerous disciplines, but its methodological implementation generally falls into two distinct categories based on the research objective: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). **Exploratory Factor Analysis (EFA)** is used when the researcher has little or no prior knowledge about the underlying structure of the data and aims to discover the number of factors and the specific variables associated with each factor. EFA is particularly useful in the early stages of test construction or theory development, allowing the data structure to dictate the latent constructs, adhering to the principle of "letting the data speak."

In contrast, **Confirmatory Factor Analysis (CFA)** is a powerful, theory-driven statistical technique. CFA requires the researcher to explicitly specify the hypothesized relationships between the observed variables and the latent factors based on established theory or prior research. The researcher must define the exact number of factors, which variables load onto which factors, and whether the factors are correlated. CFA then tests how well this pre-specified model fits the observed data, providing statistical indices (e.g., Chi-square, RMSEA, CFI) to evaluate model adequacy. CFA is a cornerstone of instrument validation, used widely to confirm the construct validity and reliability of psychological scales, surveys, and clinical assessments.

Beyond psychometrics, Factor Analysis has broad utility. In **Marketing Research**, FA is used to reduce complex consumer preference data into core dimensions (e.g., identifying underlying factors of brand loyalty or purchase intention). In **Economics** and Finance, dynamic factor models are used to distill large panels of economic indicators (like employment rates, inflation, and GDP) into a few key indices that drive macroeconomic trends. Similarly, in Genetics, FA helps identify common genetic factors that underlie complex traits or diseases. The common thread across these applications is the efficient modeling of complex, observed variability by attributing it to a simpler, latent structure, thereby enhancing prediction and theoretical understanding.

## 7. Assumptions, Limitations, and Criticisms

For Factor Analysis to yield reliable and interpretable results, several key statistical assumptions must be met, though the robustness of the technique to violations varies. The data should ideally be measured at the interval or ratio level, although ordinal data is frequently used, particularly in the social sciences, provided the number of scale points is sufficiently large. A core assumption, particularly for inferential methods like Maximum Likelihood FA, is that the variables follow a **multivariate normal distribution**. Although other methods (like PAF) are less reliant on strict normality, extreme deviations can bias the standard errors and model fit statistics.

Perhaps the most critical practical requirement for Factor Analysis is **adequate sample size**. Factor structures are unstable in small samples, leading to poor generalizability. While rules of

thumb vary widely (e.g., a minimum of 100 observations, or a subject-to-variable ratio of 10:1), large samples are necessary to reliably estimate the correlation matrix and the subsequent factor loadings. Another crucial assumption is the presence of **sufficient correlation** among variables; if the observed variables are highly independent, the FA technique cannot find shared variance, and the resulting factors will be meaningless. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity are typically used to test whether the data structure is suitable for FA.

Despite its widespread use, Factor Analysis faces several significant criticisms. Critics often point out the **subjectivity inherent in interpretation**. The number of factors retained (e.g., using the scree plot) and the decision regarding the type of rotation (orthogonal vs. oblique) often rely on researcher judgment rather than strict statistical criteria. Furthermore, the very nature of latent variables makes them theoretical constructions, not directly measurable entities, leading to debates about their ontological status. Philosophical critics argue that FA merely provides a mathematically convenient summary of correlations without necessarily proving the existence of an underlying causal structure. Finally, a significant methodological challenge lies in the **non-uniqueness of factor solutions**; different extraction methods or rotation choices can lead to statistically equivalent solutions that have vastly different theoretical interpretations, underscoring the need for strong theoretical justification throughout the process.

## Further Reading

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

[Psychometrics](#)

[Dimensionality Reduction](#)

[Charles Spearman](#)

[Louis L. Thurstone](#)