Short summary of Factor Analysis as applied in this study:

Factor analysis is proposed to reveal unobserved factors explaining variation of data. Multiple linear regression models can be stated for $n$ unobserved factors and $p$ observable variables. The linear regression model is a linear function of $n$ unobserved (common) factors plus an error ($\varepsilon$, specific factor). Hence general form of the model in matrix notation should be expressed in equation:

$$X = \mu + Lf + \varepsilon$$

Where $X$ is vector of traits with $X_i$ denoting as observable trait

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

Term $\mu$ is a vector of population means of variable $I$; $\mu_i$, represent an intercept in terms of linear regression lines.

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

$f$ is a common factor vector of $f_i$:

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$$

The number of factors $n$ should be less than $p$ ($n<p$; in our case: $n=3$ (after eigenvalue estimation, see below) and $p=6$).

$L$ is matrix of factor loadings:

$$L = \begin{pmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & \ddots & \vdots \\ l_{p1} & \cdots & l_{pn} \end{pmatrix}$$

The term $\varepsilon_i$ is a specific factor for variable $i$, incorporated into a vector of specific factors

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

Factor analysis relies on the following basic assumptions:

1. Specific and common factors are uncorrelated
2. Mean of specific factors is zero
3. Mean of common factors is zero
4. Variance of common factor is one
5. Specific factors have specific variances ($\theta_i$)
The outcome of the model and assumption is:

Variance of trait i:

\[ \sigma^2 = \text{var}(X_i) = \sum_{j=1}^{n} l_{ij}^2 + \theta_i \]

Where \( \sum_{j=1}^{n} l_{ij}^2 \) is communality for variable \( i \), sum of the squared loadings.

Covariance between pairs of traits \( i \) and \( j \) is:

\[ \sigma_{ij} = \text{cov}(X_i, X_j) = \sum_{k=1}^{n} l_{ik}l_{jk} \]

Covariance between trait \( i \) and factor \( j \) is factor loading \( \text{cov}(X_i, f_j) = l_{ij} \).

Hence, the model will be shaped in variance-covariance matrix form as:

\[ \Sigma = LL' + \epsilon \]

Several extraction methods are exploited for factor extraction. We used two methods as follows.

**Principal component extraction method**

For \( i \)-th object vector of observation with \( p \)-variables will be defined as vector:

\[ X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix} \]

Hence, the sample variance-covariance matrix will be:

\[ S = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{x})(X_i - \bar{x})' \]

(In our case, \( m \) was in range of 408 to 520, depending on analytical replicates) Hence we should have \( p \) eigenvalues \( (\lambda) \) and eigenvectors \( (\theta) \). Eigenvalue for a given factor is measurement of the variance in the entire variable which is accounted for by that factor. Thus, if a factor has a low eigenvalue its contribution to the variable can be ignored.

Our variance-covariance matrix can be expressed as the sum of \( p \)-eigenvalues multiplied by their eigenvectors and their transposes:

\[ \Sigma = \sum_{i=1}^{p} \lambda_i \theta_i \theta_i' = \sum_{i=1}^{n} \lambda_i \theta_i \theta_i' \]

The estimator of the factor loading is:

\[ \hat{l}_{ij} = \hat{\theta}_{ij} / \sqrt{\lambda_i} \]
Communalities (squared multiple correlations) are an additional parameter, which estimates reliability of factor analysis and are computed as sum of the squared loadings for the corresponding variable:

\[ \hat{h}_i = \sum_{j=1}^{n} \hat{l}_{ij}^2 \]

Communality measures the percent of variance in a given variable explained by all the factors jointly and hence it may be used as a reliability indicator.

Communality enables easy calculation of specific variances after data standardization

\[ \hat{\sigma}_i^2 = 1 - \hat{h}_i \]

Under principal component extraction method the estimated factor loadings do not change as the number of factors increase, however communalities and specific variances depend on the number of factors. As the number of factors increases, communals move towards one and the specific error falls to zero. This creates a conflicting situation wherein minimization of the number of factors is desired, however it increases residual error.

In addition, principal component based extraction does not give any information about goodness of fit.

**Maximum Likelihood Estimation method**

This method assumes that data are independently sampled from a multivariate normal distribution with \( \mu \) mean vector and the variance-covariance matrix will be expressed as:

\[ \Sigma = LL' + \Theta \]

Estimator for the population mean vector, the factor loading and specific variances are obtained by maximization the log-likelihood by the following expression:

\[ L(\mu, L, \Theta) = -\frac{mp}{2} \log 2\pi - \frac{m}{2} \log |LL'| + \Theta| - \frac{1}{2} \sum_{i=1}^{m} (X_i - \mu)' (LL' + \Theta)^{-1} (X_i - \mu) \]

The approach of the maximal likelihood is designed to find values for \( \mu, L \) and \( \Theta \) that will be the most compatible with the observed data.

The drawback of this method is that it assumes normal multivariate distribution, however it allows goodness-of-fit evaluation. The goodness of fit test compares variance-covariance matrix under the model to random variance-covariance matrices. For goodness of fit evaluation Barlett-Corrected Likelihood ratio test can be used:

\[ X^2 = (m - 1 - \frac{2p+4n-5}{6}) \ln \left( \frac{|LL'| + \Theta}{|\Sigma|} \right) \]

Under null hypothesis factor model adequately describes relationship among the variables. The hypothesis is rejected when \( X^2 \) is above critical value of chi-square. (In our experiments \( X^2 \) was in very low range from 0.00004 to 0.008).
**Factor rotations**

It is sometimes difficult to interpret factors based on loadings values alone. Rotation clarifies the output and facilitates factor interpretation. Rotation does not affect the sum of eigenvalues, but it alters eigenvalues of particular factor and will change the factor loadings. Since rotation affects factor interpretation, different rotations should be tested to find the optimal one for factor interpretation.

Orthogonal rotations do not produce factor correlation matrices as the correlation of any factor is zero.

Varimax rotation is orthogonal, which maximizes the variance of the squared loadings of a factor on all the variables in a factor matrix. It is the most common rotation method.

\[ V = \frac{1}{p} \sum_{j=1}^{n} \left( \sum_{i=1}^{p} (l_{ij})^4 - \frac{1}{p} \left( \sum_{i=1}^{p} (l_{ij})^2 \right)^2 \right) \]

Quatimix rotation is orthogonal minimizing the number of factors needed to explain each variable.

\[ Q = \frac{1}{p} \sum_{i=1}^{n} \sum_{j=1}^{p} (l_{ij})^4 \]

Equimix rotation is balance between varimax and quartimax

\[ E = \frac{1}{p} \sum_{j=1}^{n} \left( \sum_{i=1}^{p} (l_{ij})^4 - \frac{n}{2p} \left( \sum_{i=1}^{p} (l_{ij})^2 \right)^2 \right) \]

Direct oblimin rotation is a non-orthogonal method, which allows factors to be correlated. It increases eigenvalue, but decreases interpretability.

**Factor scores**

Factor scores are scores of each case on each factor. Factor scores are calculated by multiplication of the case’s standardized score by the factor loadings. Factor score allows data extraction, factor outlier analysis, data modeling and etc.

Per case level factor scores can be evaluated by the equation:

\[ Y = \mu + Lf_i + \varepsilon \]

Several methods allow factor score resolution, and the least squares method is one of the most commonly used.