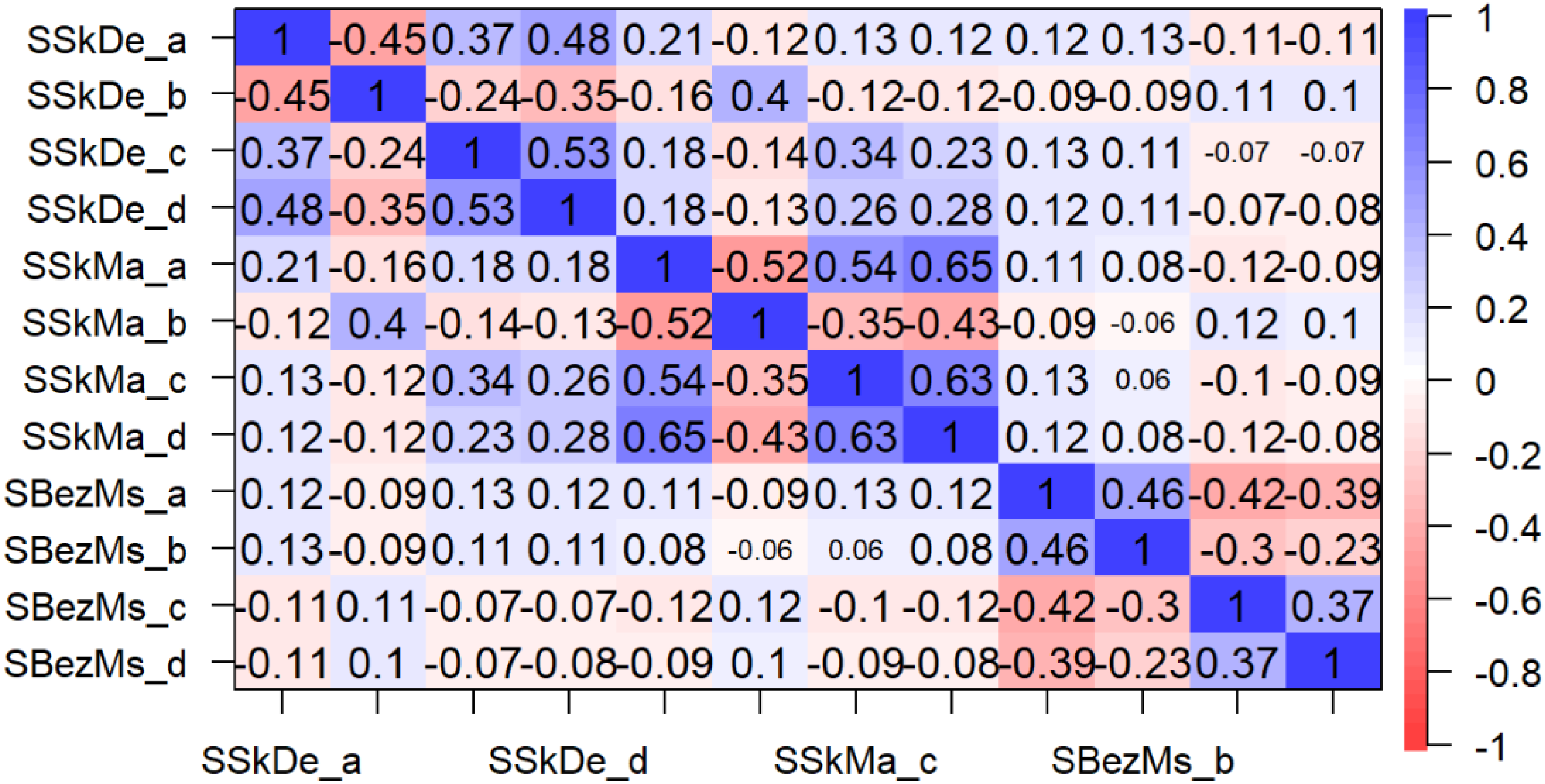
Exploratory factor analysis (EFA)

Albert-Ludwigs-Universität Freiburg

15.02.2022

Julius Fenn, M.Sc.

Correlation plot



A sizable number of correlations should exceed ±.30

Literature

- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219-246, download: https://journals.sagepub.com/doi/pdf/10.1177/0095798418771807
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, *14*(1), 20, download: https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1226&context=pare
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.
- Mvududu, N. H., & Sink, C. A. (2013). Factor analysis in counseling research and practice. *Counseling Outcome Research and Evaluation*, *4*(2), 75-98.
- …
- *German*: Chapter 23 in Moosbrugger, H., & Kelava, A. (2020). Testtheorie und Fragebogenkonstruktion, download: https://link.springer.com/book/10.1007/978-3-662-61532-4

Background

Introduction

- The term "**factor analysis**" covers a group of multivariate analysis procedures that examine underlying common dimensions of sets of variables.

- The procedures are used for two main purposes:
  - for data reduction, by reducing the variation of a large number of variables to a significantly smaller number of common dimensions that can be used for further analyses, and
  - to examine the structure of the data (e.g. whether the questionnaire items are based more on 2 or 3 factors).

Introduction

- Three different types of procedures can be distinguished:

1. principal component analysis (PCA)
2. exploratory factor analysis (EFA)
3. confirmatory factor analysis (CFA)

Introduction: Difference between PCA and EFA I

- PCA = Principal Component Analysis

- Possible question:
  - Can the multivariate data of an entrance test be represented by a smaller number of dimensions, each of which can be determined as a weighted linear combination of the variables collected?
  - Which of the variables contribute most to the individual dimensions?

- Assumption: All variables collected are free of measurement error
- Objective: To represent the essential information from several intercorrelated variables through a smaller number of new (artificial) variables, the main principal components
- => Variance maximization

Introduction: Difference between PCA and EFA II

- PCA = Principal Component Analysis
- Advantage over EFA: No model assumptions necessary.
- Disadvantage: PCA does **not fit a statistical model to the data** with the aim of aim of constructing an "explanatory model", like EFA.
- This is why PCA is often referred to in the statistical literature not as a "factor analysis", but as an independent procedure.

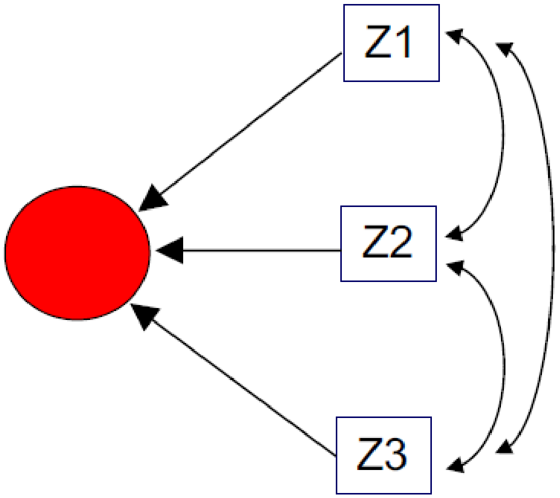
Introduction: Difference between PCA and EFA III

- EFA = Exploratory Factor Analysis
- Objective: Explanation of the structure in the correlation matrix by means of an explanatory model
  - The covariances or correlations should be explained as well as possible by the model explained by the model
  - => covariance maximization
- Advantages over PCA:
  - The relationships between the observed variables are attributed to latent variables
  - Measurement errors are taken into account
- Disadvantage: Model assumptions are necessary (e.g. normal distribution)

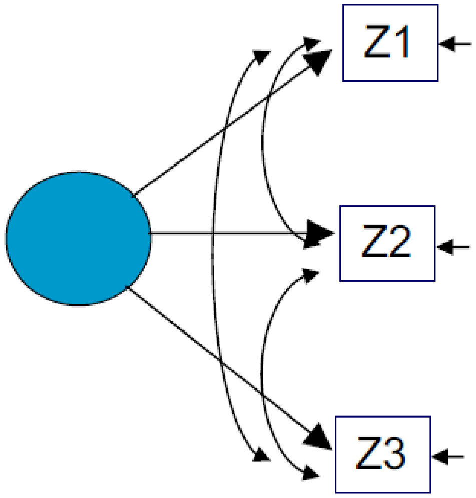
Introduction: Difference between PCA and EFA IV

- **PCA:** The main component is composed additive from the weighted items together -> all variance
- **EFA:** The correlations (covariances) between the items can be attributed to a latent variable (factor) -> common variance

PCA



EFA



Introduction: Difference between PCA and EFA V

- The basic equation of the factor analytic model is:

$$z_i = \lambda_{i,1} \cdot \xi_1 + \lambda_{i,2} \cdot \xi_2 + ... + \lambda_{i,j} \cdot \xi_j + \epsilon_i$$

- The basic equation of a PCA is:

$$z_i = \lambda_{i,1} \cdot \xi_1 + \lambda_{i,2} \cdot \xi_2 + ... + \lambda_{i,j} \cdot \xi_j$$

- Differences:
  - In an EFA, a residual variable ε is assumed:
  - That part of a manifest variable that **cannot** be attributed to the cannot be attributed to the latent variables (specific variance)
  - In an EFA, not all factors ξ are determined, but rather only a predefined number

Factor analysis

- Most of what is mentioned in the context of principal component analysis (PCA) applies conceptually in exactly the same way to exploratory factor analysis (EFA), i.e.
  - Eigenvalues and communalities
  - Extraction criteria
  - Rotation procedure

  are largely identical.
- The central difference between EFA and PCA lies in the way in which the factors (as opposed to the components) are defined

Factor analysis: Model I (! EFA)

- The basic equation of the factor analytic model is:

$$z_i = \lambda_{i,1} \cdot \xi_1 + \lambda_{i,2} \cdot \xi_2 + \ldots + \lambda_{i,j} \cdot \xi_j + \epsilon_i$$

- In matrix notation:

$$\mathbf{Z} = \Lambda \xi + \epsilon$$

$$\begin{pmatrix} z_1 \\ z_2 \\ \cdots \\ z_i \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \ldots & \lambda_{1,j} \\ \lambda_{2,1} & \lambda_{2,2} & \ldots & \lambda_{2,j} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{i,1} & \lambda_{i,2} & \ldots & \lambda_{i,j} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_j \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_i \end{pmatrix}$$

- ξ("ksi") are the factors
- Λ ("lambda") is the loading matrix with loading λ
- ε ("epsilon") is the error

Factor analysis: Model II (! EFA)

- The basic equation of the factor analytic model is:

$$z_i = \lambda_{i,1} \cdot \xi_1 + \lambda_{i,2} \cdot \xi_2 + ... + \lambda_{i,j} \cdot \xi_j + \epsilon_i$$

- The estimation is done under the following side assumptions:
  - All errors (ε) are uncorrelated
  - All (unrotated) factors (ξ) are uncorrelated
  - Errors and factors are uncorrelated
  - Factors have a mean = 0 and a variance = 1 (*this does not have to be assumed, but simplifies the presentation*)

Factor analysis: Model III (! EFA)

- FA decompose the variance of the manifest variables in:
- Total variance = **Common variance + Specific variance**

$$\sigma_i^2 = \sum_{j=1}^{J} \lambda_{i,j}^2 + \theta_i$$

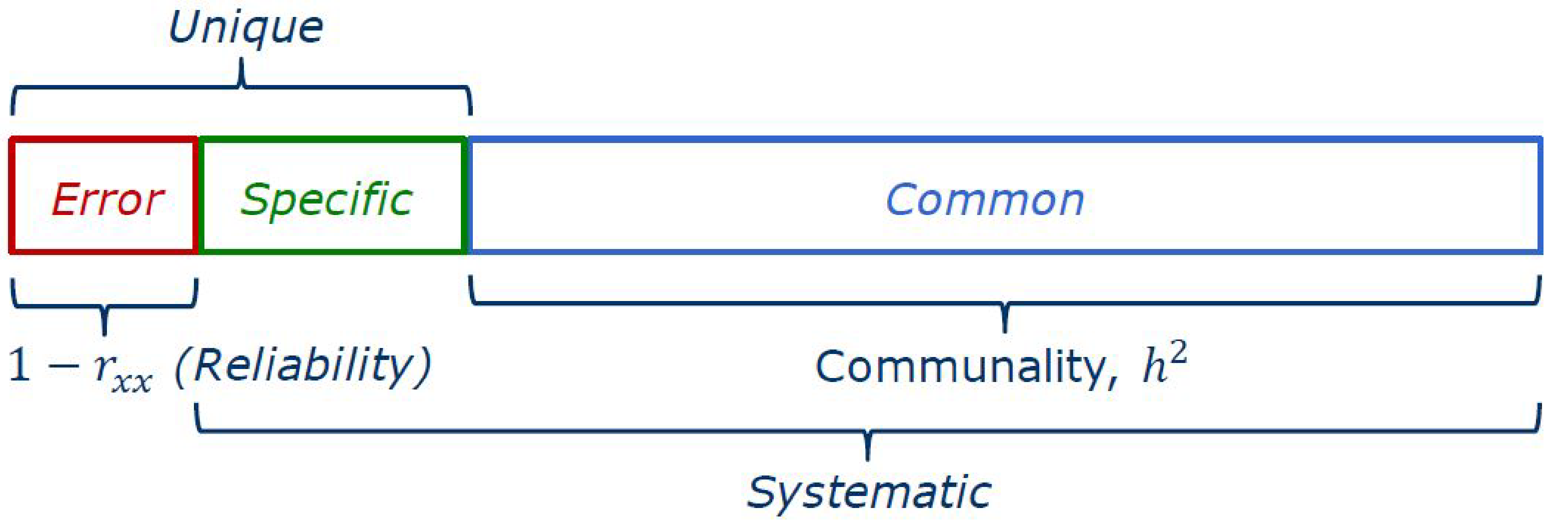
- So that the variance-covariance matrix (for one factor and three items):

$$\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \end{array} \begin{pmatrix} z_1 & z_2 & z_3 \\ \lambda_1^2 + \theta_{1,1} & & \\ \lambda_1 \cdot \lambda_2 & \lambda_2^2 + \theta_{2,2} & \\ \lambda_1 \cdot \lambda_3 & \lambda_2 \cdot \lambda_3 & \lambda_3^2 + \theta_{3,3} \end{pmatrix}$$

- => covariances between the manifest variables should be be due exclusively to common variance (factors)

Variance decomposition (! EFA)

- Within the framework of factor-analytical models (common factor model), the variance of indicators is divided into common and unique variance:



- The variance shared between the indicators is the commonality; the remaining variance is the unique variance, which is divided into indicator-specific method variance (specific) and measurement error variance (error).

Step by step

Factor analysis

- *EFA is essentially used (1) for data reduction, by reducing the variation of a large number of variables (e.g. questionnaire items) to a significantly smaller number of common dimensions that can be used for further analyses and (2) for checking the construct validity of questionnaires and tests. For example, it can be checked whether a construct is one-dimensional (e.g. Moosbrugger, 2020). Unidimensionality means that all measured items load on one factor, where one factor represents a latent construct, such as "collectivism". A variety of statistical procedures and additional tests are available for the EFA, so it is necessary to describe the individual aspects of the procedure in the following. These include (1) preliminary tests, (2) the choice of the extraction method, (3) the choice of a termination criterion and finally (4) the choice of the rotation method:*

(1) preliminary tests

- The preliminary tests (1) check whether an EFA can be applied at all. Two different tests are suitable for this: **Bartlett's test** checks whether the variables are correlated with each other. Here the null hypothesis assumes that the correlation matrix corresponds to a unit matrix and thus there are no correlations between the questions. A significant test, on the other hand, indicates ($p < .05$) that the correlation matrix is factorable, since there are dependencies between the questions. However, Bartlett's test is particularly susceptible to non-normally distributed data and should therefore be interpreted with caution. The **Kaiser-Meyer-Olkin criterion** is based on the level of variable intercorrelations and should be at least greater than .5 (e.g. Mvududu & Sink, 2013; Wolf & Best, 2010).

(2) the choice of the extraction method

- The statistical procedure for extracting the factors (2) is the **principal axis analysis** ("PAF"), which, in contrast to the principal component analysis, only analyses the common variance of the items and takes measurement errors into account. According to Moosbrugger and Schermelleh-Engel, the factors extracted in this way are actually latent variables that explain the response behavior to the underlying questions (Moosbrugger & Kelava, 2012). Furthermore, the principal axis analysis is less biased if the item responses deviate strongly from a normal distribution (Costello & Osborne, 2005)

(3) the choice of a termination criterion

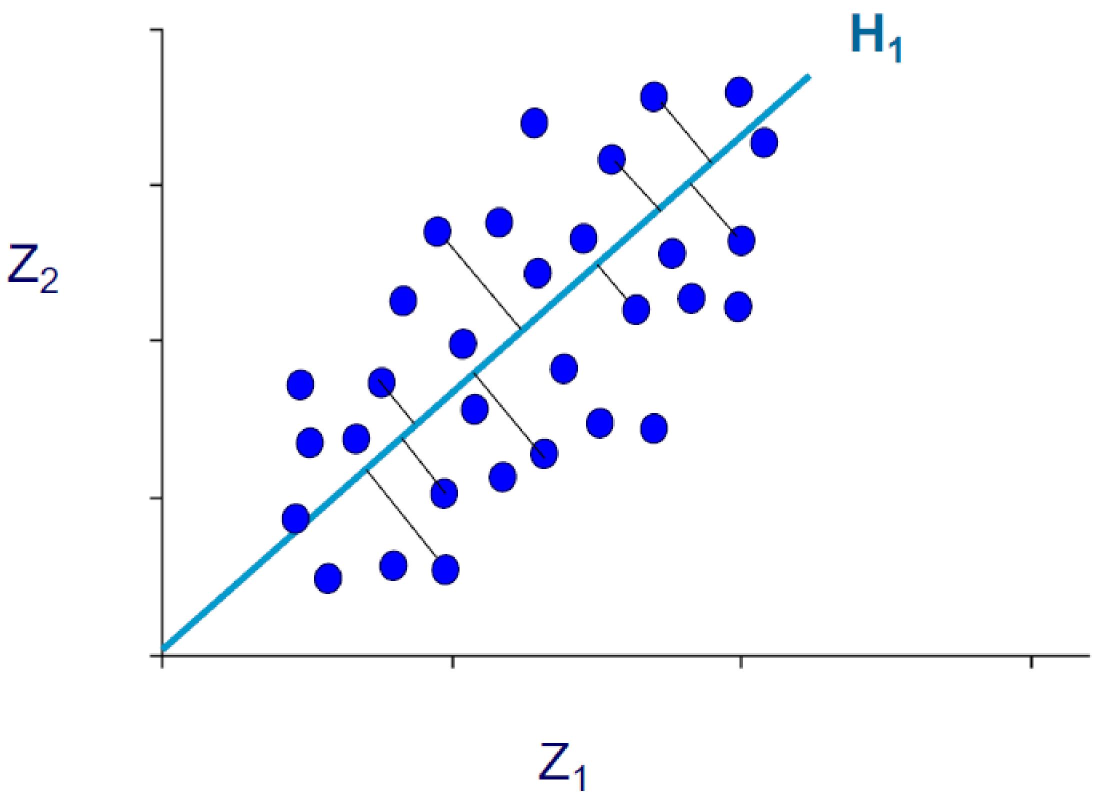
- The termination criterion (3) determines the number of factors. The basis termination criterion are the so-called "**eigenvalues**". These explain how much variance a factor explains. Here, a factor with an eigenvalue greater than one explains more than itself (e.g. an eigenvalue of 2.5 explains as much as 2.5 variables). Dividing the eigenvalues by the number of variables used in the EFA yields the explained variance share of the factor (e.g. if the first factor has an eigenvalue of 2.5 and 5 variables were used, the first factor explains 50% of the variance). For the following analyses, the **Kaiser criterion** is used, which includes all factors that have an eigenvalue greater than one (Wolf & Best, 2010). It is important to emphasize that the Kaiser criterion often overestimates the number of factors to retain and is one of the "least accurate methods for selecting the numbers of factors to retain" (Costello & Osborne, 2005, p. 2). For an initial analysis, however, the criterion is sufficient and further alternative tests of the termination criteria, such as **scree plots, parallel analysis, Velicer's MAP criterion, as well as the calculation of multiple EFAs** can be used (see e.g. Costello & Osborne, 2005b; Mvududu & Sink, 2013).

(4) the choice of the rotation method

- The aim of the last step (4) of the rotation is to interpret the content of the factors. The factors are rotated in such a way that the criterion of "**simple structure**" is fulfilled. This criterion states that each variable included in the EFA loads as highly as possible on one factor and has low loadings on the other factors (cross-loading). Here, the loadings are so-called factor loadings, which are standardized to the value range -1 to 1, whereby higher loadings indicate a stronger impact of the factor on the question (Wolf & Best, 2010). Here, a factor fulfils the criterion of "single structure" if the factor loadings are greater than +/-.4 and there are no or only minor cross-loadings on other factors. In addition, at least three variables should load on each factor, otherwise it is not statistically testable and difficult to interpret in terms of content (Costello & Osborne, 2005b; Mvududu & Sink, 2013). For the rotation of the factors, an **orthogonal rotat**ion procedure can be used, which assumes that only low correlations exist between the extracted factors. However, if factors are highly correlated with each other, this speaks in favour of a **oblique rotation**, which explicitly allows correlations between the extracted factors (Wolf & Best, 2010). In addition to the interpretation of the content of the factors and the test for one-dimensionality, the so-called "**communalities**" can be used to check how much variance of a question is explained by the extracted factors. The communalities range from 0 to 1, whereby a value closer to one means that the extracted factors can explain more variance (Mvududu & Sink, 2013).

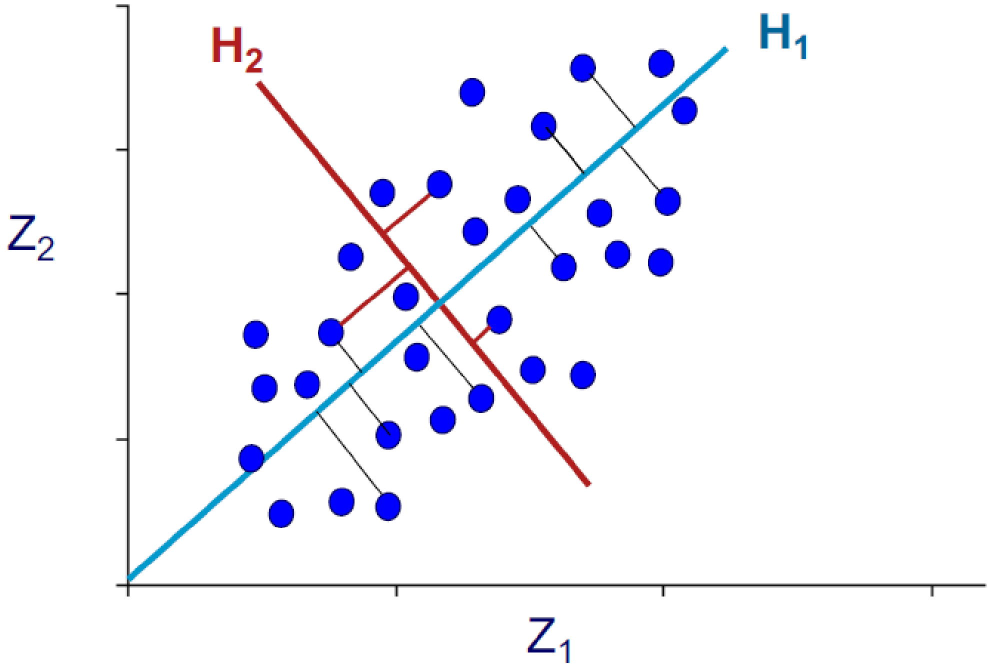
choice of the rotation method: orthogonal

- The manifestations of the persons on Z1 and Z2 scores are projected onto a new dimension (principal component)
- => The 1st principal component (H1) explains a maximum share of the total variance of the two variables.



choice of the rotation method: orthogonal

- Extracting the other main components
- => A 2nd principal component H2, orthogonal to H1, is extracted that explains as much of the unexplained variance as possible: Corr(H1,H2) = 0



Computing eigenvalues - referring to (3) (PCA) I

- The linear combinations of the principal components are determined in such a way that
  - the 1st principal component explains the maximum proportion of the total variance,
  - the 2nd principal component explains the next largest proportion of the total variance
  - all other principal components explain as much of the residual variance as possible until the entire variance is explained.
- With 2 variables, 2 principal components can be extracted, resulting in the following equations:

$$
\begin{aligned}
H_1 &= \gamma_{11} Z_1 + \gamma_{12} Z_2 \\
H_2 &= \gamma_{21} Z_1 + \gamma_{22} Z_2
\end{aligned}
$$

- Where:
  - Z = observed standardized variable (item i) H = principal component
  - *yij* = regression weight
  - m = number of items (i = 1, ..., m; here: 2)
  - p = number of principal components (j = 1, ..., p ; here: 2)
- The regression weights are to be determined in such a way that the variance of H1 is maximal

Computing eigenvalues - referring to (3) (PCA) II

- After standardisation of the principal components, the equations for the items Zi with the component loadings λij finally result:

$$
\boxed{\begin{aligned} Z_1 &= \lambda_{11} H_1 + \lambda_{12} H_2 \\ Z_2 &= \lambda_{21} H_1 + \lambda_{22} H_2 \end{aligned}}
$$

- Where:
  - …
  - λij = correlation of variable Zi with principal component Hj

- The items each represent a linear combination of the principal components.
- There are **no residual variables** in either equation!
- The differences in the observed variables are perfectly attributed to differences in the principal components

Computing eigenvalues - referring to (3) (PCA) III

- First, as many principal components are formed as there are manifest variables
  - All principal components are orthogonal to each other, they only explain the non-redundant scatter
  - The determination of the influence weights γij can be reduced to the so-called **eigenvalue problem**
  - As a constraint, the sum of the squared weights is set to 1:

$$\gamma_j'\gamma_j = 1.0 \quad (\gamma_j' = \gamma_{11} + \gamma_{12} + \ldots + \gamma_{1m})$$

$$\begin{aligned} H_1 &= \gamma_{11} Z_1 + \gamma_{12} Z_2 \\ H_2 &= \gamma_{21} Z_1 + \gamma_{22} Z_2 \end{aligned} \quad \text{->} \quad \begin{aligned} \gamma_{11}^2 + \gamma_{12}^2 &= 1.0 \\ \gamma_{21}^2 + \gamma_{22}^2 &= 1.0 \end{aligned}$$

- For each square matrix, e.g. a correlation matrix R of the type p x p, there are there are exactly p eigenvalues and p associated eigenvectors

Computing eigenvalues - referring to (3) (PCA) IV

- For every square matrix R, scalars (numbers) θ and vectors y (which do not contain only zeros) can be found for which applies:

$$\mathbf{R} \cdot \gamma = \theta \cdot \gamma$$

$$=> \mathbf{R} \cdot \gamma - \theta \cdot \gamma = 0$$

$$=> (\mathbf{R} - \theta \cdot \mathbf{I})\gamma = 0$$

- Where
  - y is the eigenvalue vector (vector of influence weights)
  - θ is the eigenvalue
  - I is the unit matrix
- The eigenvalues θ correspond to the variances of the principal components: Var(H1) = q1
  - to obtain the eigenvalues, the "characteristic equation" is solved:

$$\left|\left(\mathbf{R} - \theta \cdot \mathbf{I}\right)\right| = 0$$

  - Multiplying out leads to a quadratic equation, which is solved, and with which the eigenvalues and the vectors of the influence weights y can be determined

eigenvalue problem

- Thus, the equation to be solved is:

$$\mathbf{R}\gamma = \theta\gamma$$

  - This equation can be transformed as follows:
  - Using y = Iy and pre-brackets of y results in:

(where I is the unit matrix)

$$(\mathbf{R} - \theta \cdot \mathbf{I})\gamma = 0$$

$$\mathbf{R}\gamma - \theta\gamma = 0$$

$$(\mathbf{R} - \theta \cdot \mathbf{I})\gamma = \left(\begin{bmatrix} 1.0 & r_{12} \\ r_{21} & 1.0 \end{bmatrix} - \theta \cdot \begin{bmatrix} 1.0 & 0 \\ 0 & 1.0 \end{bmatrix} \right) \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1.0 - \theta & r_{12} \\ r_{21} & 1.0 - \theta \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Unknowns in this equation are θ and y1 and y2!

If the matrix R is positive definite, if its eigenvalues are > 0.

Central terms

Basic concepts of PCA / EFA I: eigenvalue

- The proportion of the total variance of all variables that is explained by a single principal component is called the **eigenvalue**.

- The total variance of the variables corresponds to the trace of the correlation matrix and thus the sum of the diagonal elements:

$$\mathbf{R} = \begin{bmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{bmatrix}$$

$$tr(\mathbf{R}) = 1.0 + 1.0 = 2.0$$

- Each principal component / factor has a variance, so for p principal components there are there are exactly p eigenvalues: θ 1, ..., θ p
- \- To each eigenvalue θ j belongs an eigenvector yj with weighting coefficients λij, which can be interpreted in standardized form as correlations
- \- The sum of the squared loadings of all variables on a component yields the explained variance of this component in the total variance, the eigenvalue

Basic concepts of PCA / EFA II: commonality

- **Commonality**
- A single component loading λ_{ij} can be interpreted as the correlation of the observed variable Z_i with the component H_j
- The sum of the squared charges of a variable on all the components yields the variance of this variable, which is explained by the components jointly
- This quantity is called the commonality (h_i^2) of a variable Z_i

$$h_i^2 = \sum_{j=1}^{p} \lambda_{ij}^2$$

Basic concepts of PCA / EFA III: eigenvalue and commonality

- Since standardized variables have a variance of one, an eigenvalue greater than one means that the component Hj explains more variance than a single variable Zi has
  - Only in this case does the principal component fulfil the data-reducing purpose of PCA

- Since the observed variables are usually z-standardized , the communality determined for a variable can be at most 1.0 in the case of PCA.
- In the case of EFA, however, the communality can only be as large as the reliability of the respective variable. reliability of the respective variable

Basic concepts of PCA / EFA IV: eigenvalue and commonality

UNI FREIBURG

EFA (oblimin), $r = .27$

	L1	L2	h^2
i1	.78	.05	.64
i2	.86	-.05	.71
i3	.58	-.02	.35
i4	.02	.75	.57
i5	.00	.75	.56
i6	-.04	.64	.64
% Var	28	26	

PCA (oblimin), $r = .21$

	L1	L2	h^2
i1	.85	.06	.74
i2	.88	-.04	.77
i3	.77	-.02	.59
i4	.04	.92	.69
i5	.02	.83	.69
i6	-.06	.80	.62
% Var	35	34	

- PCAs are better at explaining variance but worse at explaining covariances

Application

For applications see great examples online:

- Clark (2020, April 10). Michael Clark: Factor Analysis with the psych package. Retrieved from https://m-clark.github.io/posts/2020-04-10-psych-explained/
- See chapter 6 in https://personality-project.org/r/book/
- See ETH Zürich [german]: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/interdependenz/reduktion/faktor.html