Lecture 11: PCA vs CFA

Lecturer: Ksenia Kasianova
xeniakasianova@gmail.com

February 5, 2024

Plan

1) PCA

2) PCA vs CFA

3) CFA

Intro

Multivariate Statistics

- a group of statistical methods that focus on studying multiple variables together while focusing on the variation that those variables have in common.

Intro

Multivariate Statistics

- a group of statistical methods that focus on studying multiple variables together while focusing on the variation that those variables have in common.

- treatment of data sets with a large number of dimensions

Intro

Multivariate Statistics

- a group of statistical methods that focus on studying multiple variables together while focusing on the variation that those variables have in common.

- treatment of data sets with a large number of dimensions

- reducing the number of variables

Intro

Multivariate Statistics

- a group of statistical methods that focus on studying multiple variables together while focusing on the variation that those variables have in common.

- treatment of data sets with a large number of dimensions

- reducing the number of variables

PCA, **for Principal Component Analysis**, and **CFA, Common Factor Analysis**, are two statistical methods that are often covered together in classes on Multivariate Statistics.

PCA — Principal Component Analysis

Goal: Reducing the number of dimensions of a data set

Idea:

– When a dataset contains a large number of variables, there is often a serious amount of overlap between those variables.

PCA — Principal Component Analysis

Goal: Reducing the number of dimensions of a data set

Idea:

– When a dataset contains a large number of variables, there is often a serious amount of overlap between those variables.

– The components that are found by PCA are ordered from the highest information content to the lowest information content.

PCA — Principal Component Analysis

Goal: Reducing the number of dimensions of a data set

Idea:

– When a dataset contains a large number of variables, there is often a serious amount of overlap between those variables.

– The components that are found by PCA are ordered from the highest information content to the lowest information content.

=> "regrouping" the variables into a smaller number of variables, called components based on variation common to multiple variables:

– the first (newly created) component contains a maximum of variation

– the second component contains the second-largest amount of variation, etc.

– the last component logically contains the smallest amount of variation.

Choose only few of the newly created components rather than the original variables, while still retaining a maximum amount of variation.

PCA — Principal Component Analysis

Goal: Reducing the number of dimensions of a data set

Idea:

– When a dataset contains a large number of variables, there is often a serious amount of overlap between those variables.

– The components that are found by PCA are ordered from the highest information content to the lowest information content.

=> "regrouping" the variables into a smaller number of variables, called components based on variation common to multiple variables:

– the first (newly created) component contains a maximum of variation

– the second component contains the second-largest amount of variation, etc.

– the last component logically contains the smallest amount of variation.

Choose only few of the newly created components rather than the original variables, while still retaining a maximum amount of variation.

Use:

– data exploration, finding patterns in data of high dimension

– face recognition and image compression

PCA — Principal Component Analysis

Mathematical Model - Maximize variance of the new components

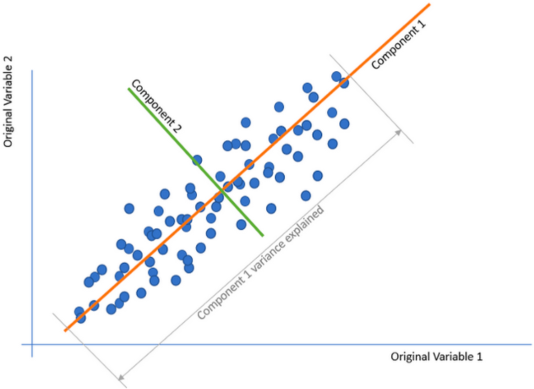


Figure: Schematic model of PCA

Mathematical definition of the PCA problem: find a linear combination of the original variables with maximum variance.

Mathematical Model - Maximize variance of the new components

Mathematical Model - Maximize variance of the new components

1) Start with a (new) component z.

z is going to be computed based on our original variables $(\mathrm{X}_1, \mathrm{X}_2, \dots)$ multiplied by a weight for each of our variables $(u_1, u_2, ...)$.

This can be written as $\mathbf{z} = \mathbf{Xu}$.

Mathematical Model - Maximize variance of the new components

1) Start with a (new) component z.

z is going to be computed based on our original variables $(\mathrm{X}_1, \mathrm{X}_2, \dots)$ multiplied by a weight for each of our variables $(u_1, u_2, ...)$.

This can be written as $\mathbf{z} = \mathbf{X}\mathbf{u}$.

2) The mathematical goal is to find the values for $\mathbf{u}$ that will maximize the variance of $\mathbf{z}$, with a constraint of unit length on $\mathbf{u}$.

Mathematical Model - Maximize variance of the new components

1) Start with a (new) component z.

z is going to be computed based on our original variables $(\mathrm{X}_1, \mathrm{X}_2, \dots)$ multiplied by a weight for each of our variables $(u_1, u_2, ...)$.

This can be written as $\mathbf{z} = \mathbf{X}\mathbf{u}$.

2) The mathematical goal is to find the values for **u** that will maximize the variance of **z**, with a constraint of unit length on **u**.

Solution:

– This problem is mathematically called **a constrained optimization using Lagrange Multiplier**

– In practice, sequential numerical optimization.

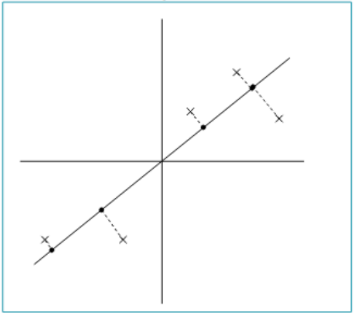
– Can be described as applying matrix decomposition to the correlation matrix of the original variables.

Goal: Maximizing the Variance

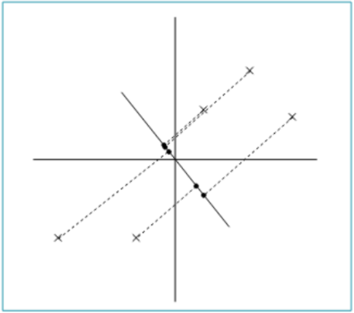
- Consider the two projections below

- Which maximizes the variance?

Option A



Option B



1) **PCA algorithm I (sequential)**

Given the centered data $\{\mathrm{x}_1, \ldots, \mathrm{x}_m\}$, compute the principal vectors:

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m} \sum_{i=1}^{m} \left\{ \left(\mathbf{w}^T \mathbf{x}_\mathrm{i} \right)^2 \right\} \quad 1^{\text{st}} \text{ PCA vector}$$

1) **PCA algorithm I (sequential)**

Given the centered data $\{\mathrm{x}_1, \ldots, \mathrm{x}_m\}$, compute the principal vectors:

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m} \sum_{i=1}^{m} \left\{ \left(\mathbf{w}^T \mathbf{x}_\mathrm{i} \right)^2 \right\} \quad 1^{\mathrm{st}} \text{ PCA vector}$$

We maximize the variance of projection of $\mathbf{x}$

$$\mathbf{w}_k = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m} \sum_{i=1}^{m} \left\{ \left[\mathbf{w}^T \left(\mathbf{x}_i - \sum_{j=1}^{k-1} \mathbf{w}_j \mathbf{w}_j^T \mathbf{x}_i \right) \right]^2 \right\} \quad k^{\mathrm{th}} \text{ PCA vector}$$

where $\sum_{j=1}^{k-1} \mathbf{w}_j \mathbf{w}_j^T \mathbf{x}_i$ is $\mathbf{x}'$ PCA reconstruction

1) **PCA algorithm I (sequential)**

Given the centered data $\{x_1, \ldots, x_m\}$, compute the principal vectors:

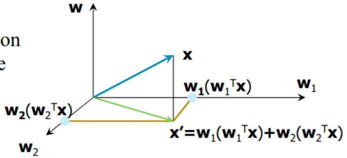
$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \frac{1}{m} \sum_{i=1}^{m} \left\{ \left(\mathbf{w}^T \mathbf{x}_i \right)^2 \right\} \quad 1^{\text{st}} \text{ PCA vector}$$

We maximize the variance of projection of $\mathbf{x}$

$$\mathbf{w}_k = \arg \max_{\|\mathbf{w}\|=1} \frac{1}{m} \sum_{i=1}^{m} \left\{ \left[\mathbf{w}^T \left(\mathbf{x}_i - \sum_{j=1}^{k-1} \mathbf{w}_j \mathbf{w}_j^T \mathbf{x}_i \right) \right]^2 \right\} \quad k^{\text{th}} \text{ PCA vector}$$

where $\sum_{j=1}^{k-1} \mathbf{w}_j \mathbf{w}_j^T \mathbf{x}_i$ is $\mathbf{x}'$ PCA reconstruction

We maximize the variance of the projection in the residual subspace



Slide from Barnabas Poczos

2) **PCA algorithm II (sample covariance matrix)**

Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_\mathrm{m}\}$, compute covariance matrix Σ

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \bar{\mathbf{x}})\,(\mathbf{x} - \bar{\mathbf{x}})^T \quad \text{where } \bar{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i$$

PCA basis vectors = the eigenvectors of Σ

Lecture 11

Ksenia Kasianova

Plan

Intro

PCA — Principal Component Analysis

CFA – Common Factor Analysis

Summary

2) **PCA algorithm II (sample covariance matrix)**

Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_\mathrm{m}\}$, compute covariance matrix Σ

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T \quad \text{where } \overline{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i$$

PCA basis vectors $=$ the eigenvectors of Σ

By finding the eigenvalues and eigenvectors of the covariance matrix, we find that

– the eigenvectors with the largest eigenvalues correspond to the dimensions that have the **strongest correlation in the dataset** – the principal components.

2) **PCA algorithm II (sample covariance matrix)**

Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_\mathrm{m}\}$, compute covariance matrix Σ

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T \quad \text{where } \overline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i$$

PCA basis vectors = the eigenvectors of Σ

By finding the eigenvalues and eigenvectors of the covariance matrix, we find that

– the eigenvectors with the largest eigenvalues correspond to the dimensions that have the **strongest correlation in the dataset** – the principal components.

– larger eigenvalue $\Rightarrow$ more important eigenvectors PCA

Power iteration (Von Mises iteration is a standard algorithm for this)

Why the Eigenvectors?

Maximise $\quad \mathbf{u}^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}$

s.t $\quad \mathbf{u}^\top \mathbf{u} = 1$

Lecture 11
Ksenia Kasianova
Plan
Intro
PCA — Principal Component Analysis
CFA – Common Factor Analysis
Summary

Why the Eigenvectors?

Maximise $\mathbf{u}^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}$

s.t $\mathbf{u}^\top \mathbf{u} = 1$

Construct Langrangian $\mathbf{u}^\top \mathbf{X}\mathbf{X}^\top \mathbf{u} - \lambda \mathbf{u}^\top \mathbf{u}$

Why the Eigenvectors?

Maximise $\mathbf{u}^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}$

s.t $\mathbf{u}^\top \mathbf{u} = 1$

Construct Langrangian $\mathbf{u}^\top \mathbf{X}\mathbf{X}^\top \mathbf{u} - \lambda \mathbf{u}^\top \mathbf{u}$

Vector of partial derivatives set to zero

$$\mathbf{x}\mathbf{x}^\top \mathbf{u} - \lambda \mathbf{u} = \left(\mathbf{x}\mathbf{x}^\top - \lambda \mathbf{I}\right) \mathbf{u} = 0$$

As $\mathbf{u} \neq \mathbf{0}$ then $\mathbf{u}$ must be an eigenvector of $\mathbf{X}\mathbf{X}^\top$ with eigenvalue λ

Eigenvalues & Eigenvectors

- For symmetric matrices, eigenvectors for distinct eigenvalues are **orthogonal**

$$Sv_{\{1,2\}} = \lambda_{\{1,2\}} v_{\{1,2\}}, \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow v_1 \bullet v_2 = 0$$

Eigenvalues & Eigenvectors

- For symmetric matrices, eigenvectors for distinct eigenvalues are **orthogonal**

$$Sv_{\{1,2\}} = \lambda_{\{1,2\}} v_{\{1,2\}}, \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow v_1 \bullet v_2 = 0$$

- All eigenvalues of a real symmetric matrix are **real**.

$$\text{if } |S - \lambda I| = 0 \text{ and } \mathrm{S} = \mathrm{S}^{\mathrm{T}} \Rightarrow \lambda \in \Re$$

Eigenvalues & Eigenvectors

- For symmetric matrices, eigenvectors for distinct eigenvalues are **orthogonal**

$$Sv_{\{1,2\}} = \lambda_{\{1,2\}} v_{\{1,2\}}, \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow v_1 \bullet v_2 = 0$$

- All eigenvalues of a real symmetric matrix are **real**.

$$\text{if } |S - \lambda I| = 0 \text{ and } \mathrm{S} = \mathrm{S}^{\mathrm{T}} \Rightarrow \lambda \in \Re$$

- All eigenvalues of a positive semidefinite matrix are **nonnegative**

$$\forall w \in \Re^n, w^T S w \geq 0, \text{ then if } \mathrm{Sv} = \lambda \mathrm{v} \Rightarrow \lambda \geq 0$$

Eigen/diagonal Decomposition

- Let $\mathbf{S} \in \mathbb{R}^{m \times m}$ be a square matrix with m linearly independent eigenvectors (a "non-defective" matrix)

Eigen/diagonal Decomposition

- Let $\mathbf{S} \in \mathbb{R}^{m \times m}$ be a square matrix with m linearly independent eigenvectors (a "non-defective" matrix)

- Theorem: Exists an eigen decomposition (unique for distinct eigen values)

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^{-1}$$

where Λ is diagonal

Eigen/diagonal Decomposition

- Let $\mathbf{S} \in \mathbb{R}^{m \times m}$ be a square matrix with m linearly independent eigenvectors (a "non-defective" matrix)

- Theorem: Exists an eigen decomposition (unique for distinct eigen values)

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^{-1}$$

where Λ is diagonal

- Columns of $\boldsymbol{U}$ are eigenvectors of S

- Diagonal elements of $\boldsymbol{\Lambda}$ are eigenvalues of $\mathbf{S}$

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_m\right), \quad \lambda_i \geq \lambda_{i+1}$$

PCA: Two Interpretations

E.g., for the first component.

I. Maximum Variance Direction: 1^{st} PC a vector v such that projection on to this vector capture maximum variance in the data (out of all possible one dimensional projections)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{v}^T\mathbf{x}_i\right)^2 = \mathbf{v}^T\mathbf{X}\mathbf{X}^T\mathbf{v}$$

PCA: Two Interpretations

E.g., for the first component.

I. Maximum Variance Direction: 1^{st} PC a vector v such that projection on to this vector capture maximum variance in the data (out of all possible one dimensional projections)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{v}^T\mathbf{x}_i\right)^2 = \mathbf{v}^T\mathbf{X}\mathbf{X}^T\mathbf{v}$$

II. Minimum Reconstruction Error: 1^{st} PC a vector v such that projection on to this vector yields minimum MSE reconstruction

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_i - \left(\mathbf{v}^T\mathbf{x}_i\right)\mathbf{v}\right\|^2$$

PCA: Two Interpretations

E.g., for the first component.

I. Maximum Variance Direction: 1^{st} PC a vector v such that projection on to this vector capture maximum variance in the data (out of all possible one dimensional projections)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{v}^T\mathbf{x}_i\right)^2 = \mathbf{v}^T\mathbf{X}\mathbf{X}^T\mathbf{v}$$

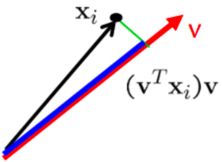
II. Minimum Reconstruction Error: 1^{st} PC a vector v such that projection on to this vector yields minimum MSE reconstruction

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_i - \left(\mathbf{v}^T\mathbf{x}_i\right)\mathbf{v}\right\|^2$$

blue ^{2}+ green ^{2} $=$ black ^{2}

black ^{2} is fixed (it's just the data)

So, maximizing blue ^{2} is equivalent to minimizing green ^{2}



3) PCA algorithm III: **SVD of the data matrix**

Singular Value Decomposition of the centered data matrix **X**.

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m] \in \mathbb{R}^{N \times m}, \quad m: \text{ number of instances }, N: \text{ number of dimensions}$$

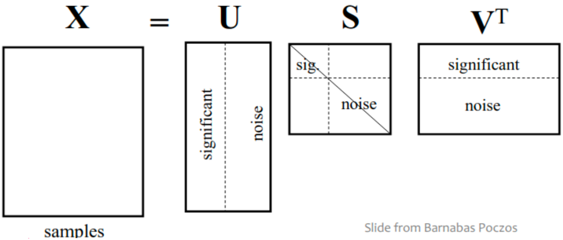
$$\mathbf{X}_{\text{features} \times \text{samples}} = \text{USV}^{\top}$$

3) PCA algorithm III: **SVD of the data matrix**

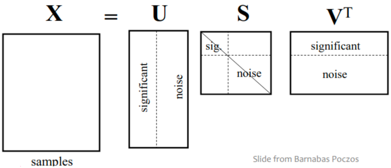
Singular Value Decomposition of the centered data matrix **X**.

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m] \in \mathbb{R}^{N\times m}, \quad m: \text{ number of instances }, N: \text{ number of dimensions}$$

$$\mathbf{X}_{\text{features } \times \text{ samples}} = \text{USV}^\top$$



Slide from Barnabas Poczos

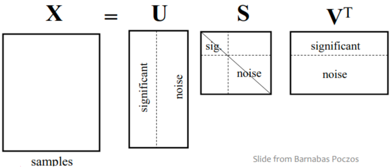


Slide from Barnabas Poczos

Columns of U

- the principal vectors, $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$
- orthogonal and has unit norm - so $U'U = I$
- Can reconstruct the data using linear combinations of $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$

PCA — Principal Component Analysis



Slide from Barnabas Poczos

Columns of U

- the principal vectors, $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$

- orthogonal and has unit norm - so $U'U = I$

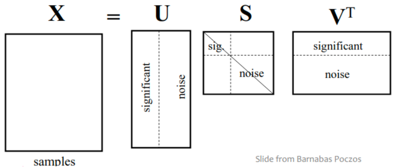
- Can reconstruct the data using linear combinations of $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$

Matrix S

- Diagonal

- Shows importance of each eigenvector

PCA — Principal Component Analysis



Columns of U

- the principal vectors, $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$

- orthogonal and has unit norm - so $U'U = I$

- Can reconstruct the data using linear combinations of $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$

Matrix S

- Diagonal

- Shows importance of each eigenvector

Columns of $\mathbf{V}^\top$

- The coefficients for reconstructing the samples

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

Factor Analysis Vs. Principle Component Analysis

- PCA components explain the maximum amount of variance while CFA explains the covariance in data.

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

Factor Analysis Vs. Principle Component Analysis

- PCA components explain the maximum amount of variance while CFA explains the covariance in data.

- PCA components are fully orthogonal to each other whereas CFA does not require factors to be orthogonal.

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

Factor Analysis Vs. Principle Component Analysis

- PCA components explain the maximum amount of variance while CFA explains the covariance in data.

- PCA components are fully orthogonal to each other whereas CFA does not require factors to be orthogonal.

- PCA component is a linear combination of the observed variable while in CFA, the observed variables are linear combinations of the unobserved variable or factor.

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

Factor Analysis Vs. Principle Component Analysis

- PCA components explain the maximum amount of variance while CFA explains the covariance in data.

- PCA components are fully orthogonal to each other whereas CFA does not require factors to be orthogonal.

- PCA component is a linear combination of the observed variable while in CFA, the observed variables are linear combinations of the unobserved variable or factor.

- PCA components are uninterpretable. In CFA, underlying factors are labelable and interpretable.

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

Factor Analysis Vs. Principle Component Analysis

- PCA components explain the maximum amount of variance while CFA explains the covariance in data.

- PCA components are fully orthogonal to each other whereas CFA does not require factors to be orthogonal.

- PCA component is a linear combination of the observed variable while in CFA, the observed variables are linear combinations of the unobserved variable or factor.

- PCA components are uninterpretable. In CFA, underlying factors are labelable and interpretable.

- PCA is a kind of dimensionality reduction method whereas factor analysis is the latent variable method.

CFA – Common Factor Analysis

PCA is efficient in finding the components that maximize variance.

Sometimes, however, we are not purely interested in maximizating variance, but to get the most useful **interpretations** to our newly defined dimensions.

Solution: Common Factor Analysis: an alternative to PCA that has a little bit more flexibility

Factor Analysis Vs. Principle Component Analysis

- PCA components explain the maximum amount of variance while CFA explains the covariance in data.

- PCA components are fully orthogonal to each other whereas CFA does not require factors to be orthogonal.

- PCA component is a linear combination of the observed variable while in CFA, the observed variables are linear combinations of the unobserved variable or factor.

- PCA components are uninterpretable. In CFA, underlying factors are labelable and interpretable.

- PCA is a kind of dimensionality reduction method whereas factor analysis is the latent variable method.

- PCA is a type of factor analysis. PCA is observational whereas FA is a modeling technique.

CFA – Common Factor Analysis

Goal - Finding latent variables in a data set

CFA allows reducing information in a larger number of variables into a smaller number of variables – "latent variables" – that make sense to us.

CFA – Common Factor Analysis

Goal - Finding latent variables in a data set

CFA allows reducing information in a larger number of variables into a smaller number of variables – "latent variables" – that make sense to us.

Unlike PCA, we can rotate the solution until we find latent variables that have a clear interpretation.

CFA – Common Factor Analysis

Goal - Finding latent variables in a data set

CFA allows reducing information in a larger number of variables into a smaller number of variables – "latent variables" – that make sense to us.

Unlike PCA, we can rotate the solution until we find latent variables that have a clear interpretation.

The principle:

– there a certain number of factors in a data set

– each of the measured variables captures a part of one or more of those factors.

Example:

There are many students in a school with grades for many subjects.

Example:

There are many students in a school with grades for many subjects.

Different grades are partly correlated: a more intellectually gifted student would have higher grades overall $=>$ a latent variable.

Example:

There are many students in a school with grades for many subjects.

Different grades are partly correlated: a more intellectually gifted student would have higher grades overall $=>$ a latent variable.

But we could also imagine having students who are overall good in languages, but bad in technical subjects.

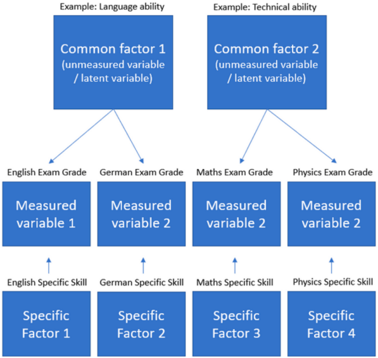
Example:

There are many students in a school with grades for many subjects.

Different grades are partly correlated: a more intellectually gifted student would have higher grades overall => a latent variable.

But we could also imagine having students who are overall good in languages, but bad in technical subjects.

In this case, we could try to find a latent variable for language ability and a second latent variable for technical ability.



CFA – Common Factor Analysis

Mathematically, CFA is similar to multiple regression analysis in that each variable is expressed as a linear combination of underlying factors.

The covariation among the variables consists of two terms (not overtly observed)

– a small number of common factors

– plus a unique factor for each variable.

CFA – Common Factor Analysis

Mathematically, CFA is similar to multiple regression analysis in that each variable is expressed as a linear combination of underlying factors.

The covariation among the variables consists of two terms (not overtly observed)

– a small number of common factors

– plus a unique factor for each variable.

If the variables are standardised, the factor model may be represented as:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \ldots + A_{im}F_m + V_iU_i$$

where $X_i = i$ th standardised variable

$A_{ij} =$ standardised multiple regression coefficient of variable i on common factor j

$F =$ common factor

$V_i =$ standardised regression coefficient of variable i on unique factor i

$U_i =$ the unique factor for variable i

$m =$ number of common factors.

Def.: Communality – the amount of variance a variable shares with all other variables included in the analysis

The factor model:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \ldots + A_{im}F_m + V_iU_i$$

where $\quad X_i = i$ th standardised variable

CFA – Common Factor Analysis

The factor model:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \ldots + A_{im}F_m + V_iU_i$$

where $X_i = i$ th standardised variable

The unique factors U_i are correlated with each other and with the common factors F.

CFA – Common Factor Analysis

The factor model:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \ldots + A_{im}F_m + V_iU_i$$

where $\quad X_i = i$ th standardised variable

The unique factors U_i are correlated with each other and with the common factors F.

The common factors themselves can be expressed as linear combinations of the observed variables:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \ldots + W_{ik}W_k$$

where $\quad F_i =$ estimate of i th factor

$W_i =$ weight or factor score coefficient

$k =$ number of variables.

CFA – Common Factor Analysis

The factor model:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \ldots + A_{im}F_m + V_iU_i$$

where $X_i = i$ th standardised variable

The unique factors U_i are correlated with each other and with the common factors F.

The common factors themselves can be expressed as linear combinations of the observed variables:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \ldots + W_{ik}W_k$$

where $F_i =$ estimate of i th factor

$W_i =$ weight or factor score coefficient

$k =$ number of variables.

It is possible to select weights or factor score coefficients so that:

– the first factor explains the largest portion of the total variance

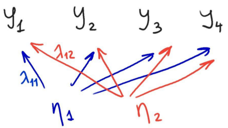
– the second factor accounts for most of the residual variance, subject to being uncorrelated with the first factor (the second highest variance), etc.

Example (cont.):

Consider latent variables: the general ability of a student for Language and Technical subjects.

Example (cont.):

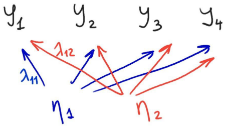
Consider latent variables: the general ability of a student for Language and Technical subjects.



CFA – Common Factor Analysis

Example (cont.):

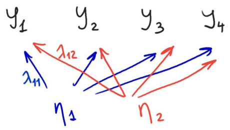
Consider latent variables: the general ability of a student for Language and Technical subjects.



Possible that some students are great at languages overall, but that they are just bad at German => specific factors which measure the impact of one variable on the measured variable.

Example (cont.):

Consider latent variables: the general ability of a student for Language and Technical subjects.



Possible that some students are great at languages overall, but that they are just bad at German => specific factors which measure the impact of one variable on the measured variable.

$$y_1 = \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \varepsilon_1$$

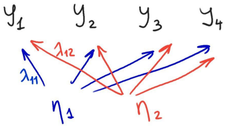
$$\vdots$$

$$y_4 = \underbrace{\lambda_{41}\eta_1 + \lambda_{42}\eta_2}_{\text{common}} + \underbrace{\varepsilon_4}_{\text{unique}}$$

CFA – Common Factor Analysis

Example (cont.):

Consider latent variables: the general ability of a student for Language and Technical subjects.



Possible that some students are great at languages overall, but that they are just bad at German => specific factors which measure the impact of one variable on the measured variable.

$$y_1 = \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \varepsilon_1$$
$$\vdots$$
$$y_4 = \lambda_{41}\eta_1 + \lambda_{42}\eta_2 + \varepsilon_4$$

common unique

– λ_{ij} shows "Ability for learning German while taking into account the general ability for learning languages".

CFA – Common Factor Analysis

Common factor model:

$$y_1 = \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \varepsilon_1$$

$$\vdots$$

$$y_4 = \lambda_{41}\eta_1 + \lambda_{42}\eta_2 + \varepsilon_4$$

common unique

Common factor model:

$$
\begin{aligned}
y_1 &= \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \varepsilon_1 \\
&\vdots \\
y_4 &= \underbrace{\lambda_{41}\eta_1 + \lambda_{42}\eta_2}_{\text{common}} + \underbrace{\varepsilon_4}_{\text{unique}}
\end{aligned}
$$

$$
\begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \vdots & \vdots \\ \lambda_{41} & \lambda_{42} \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_1 \end{pmatrix}
$$

CFA – Common Factor Analysis

Common factor model:

$$y_1 = \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \varepsilon_1$$

$$\vdots$$

$$y_4 = \lambda_{41}\eta_1 + \lambda_{42}\eta_2 + \varepsilon_4$$

common unique

$$\begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \vdots & \vdots \\ \lambda_{41} & \lambda_{42} \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_1 \end{pmatrix}$$

– λ_{i1} attributes one part of its variation to first common latent variable (η_1)

– λ_{i2} attributes one part of its variation to first common latent variable (η_2)

– part of variation related to a specific factor (specific to this variable; ϵ).

$$\begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \vdots & \vdots \\ \lambda_{41} & \lambda_{42} \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_1 \end{pmatrix}$$

$$\bar{y} = \Lambda\bar{\eta} + \bar{\varepsilon}$$

where $\Lambda\bar{\eta}$ – common fixed weighting of unobserved factors

Λ are the values that we need to estimate.

$$\begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \vdots & \vdots \\ \lambda_{41} & \lambda_{42} \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_1 \end{pmatrix}$$

$$\bar{y} = \Lambda\bar{\eta} + \bar{\varepsilon}$$

where $\Lambda\bar{\eta}$ – common fixed weighting of unobserved factors

Λ are the values that we need to estimate.

To solve this, the same mathematical solution as in PCA is used, except for a small difference.

– In PCA we apply matrix decomposition to the correlation matrix.

– In Factor Analysis, we apply matrix decomposition to a correlation matrix in which the diagonal entries are replaced by $1 - \text{var}(\mathbf{d})$, one minus the variance of the specific factor of the variable.

CFA – Common Factor Analysis

Assumptions:

1. There are no outliers in data.

CFA – Common Factor Analysis

Assumptions:

1. There are no outliers in data.
2. Sample size should be greater than the factor.

CFA – Common Factor Analysis

Assumptions:

1. There are no outliers in data.

2. Sample size should be greater than the factor.

3. The variables used in factor analysis should be linearly related to each other. This can be checked by looking at scatterplots of pairs of variables.

CFA – Common Factor Analysis

Assumptions:

1. There are no outliers in data.

2. Sample size should be greater than the factor.

3. The variables used in factor analysis should be linearly related to each other. This can be checked by looking at scatterplots of pairs of variables.

4. There should not be perfect multicollinearity.

However, the variables must also be at least moderately correlated to each other, otherwise the number of factors will be almost the same as the number of original variables, which means that carrying out a factor analysis would be pointless.

CFA – Common Factor Analysis

Assumptions:

1. There are no outliers in data.

2. Sample size should be greater than the factor.

3. The variables used in factor analysis should be linearly related to each other. This can be checked by looking at scatterplots of pairs of variables.

4. There should not be perfect multicollinearity.

However, the variables must also be at least moderately correlated to each other, otherwise the number of factors will be almost the same as the number of original variables, which means that carrying out a factor analysis would be pointless.

5. There should not be homoscedasticity between the variables.

CFA – Common Factor Analysis

Assumptions:

1. There are no outliers in data.

2. Sample size should be greater than the factor.

3. The variables used in factor analysis should be linearly related to each other. This can be checked by looking at scatterplots of pairs of variables.

4. There should not be perfect multicollinearity.

However, the variables must also be at least moderately correlated to each other, otherwise the number of factors will be almost the same as the number of original variables, which means that carrying out a factor analysis would be pointless.

5. There should not be homoscedasticity between the variables.

6. Factor analysis is designed for interval data, although it can also be used for ordinal data (e.g. scores assigned to Likert scales).

Terminology:

Factor – a latent (hidden, unobserved) variable which describes the association among the number of observed variables.

– The maximum number of factors are equal to a number of observed variables.

– Every factor explains a certain variance in observed variables.

–The factors with the lowest amount of variance were dropped.

Terminology:

Factor – a latent (hidden, unobserved) variable which describes the association among the number of observed variables.

– The maximum number of factors are equal to a number of observed variables.

– Every factor explains a certain variance in observed variables.

–The factors with the lowest amount of variance were dropped.

Eigenvalues – represent variance explained each factor from the total variance.

CFA – Common Factor Analysis

Terminology:

Factor – a latent (hidden, unobserved) variable which describes the association among the number of observed variables.

– The maximum number of factors are equal to a number of observed variables.

– Every factor explains a certain variance in observed variables.

–The factors with the lowest amount of variance were dropped.

Eigenvalues – represent variance explained each factor from the total variance.

Factor Rotation – re-distributed the commonalities with a clear pattern of loadings.

– Rotation is a tool for better interpretation of factor analysis.

– Rotation can be orthogonal or oblique.

CFA – Common Factor Analysis

Factor loading – a matrix which shows

– the relationship of each variable to the underlying factor,

– the correlation coefficient for observed variable and factor,

– the variance explained by the observed variables.

CFA – Common Factor Analysis

Factor loading – a matrix which shows

– the relationship of each variable to the underlying factor,

– the correlation coefficient for observed variable and factor,

– the variance explained by the observed variables.

Communalities – the sum of the squared loadings for each variable.

– represents the common variance

– ranges from $0 - 1$ and value close to 1 represents more variance.

CFA – Common Factor Analysis

There are three main steps in a factor analysis:

1. Calculate initial factor loadings.

– **Principal component method:** looks for a set of factors which can account for the total variability in the original variables.

– **Principal axis factoring:** tries to find the lowest number of factors which can account for the variability in the original variables that is associated with these factors

CFA – Common Factor Analysis

There are three main steps in a factor analysis:

1. Calculate initial factor loadings.

– **Principal component method:** looks for a set of factors which can account for the total variability in the original variables.

– **Principal axis factoring:** tries to find the lowest number of factors which can account for the variability in the original variables that is associated with these factors

Two methods will tend to give similar results if

- the variables are quite highly correlated

- the number of original variables is quite high.

Whichever method is used, the resulting factors at this stage will be uncorrelated.

2. Factor Rotation:

Goal: to improve the overall interpretability.

2. Factor Rotation:

Goal: to improve the overall interpretability.

If there are 'clusters' (groups) of variables that are strongly inter-related

– the rotation is done to try to make variables **within a subgroup** score as highly (positively or negatively) as possible on one particular factor while

– ensuring that the loadings for these variables on the remaining factors are as low as possible.

2. Factor Rotation:

Goal: to improve the overall interpretability.

If there are 'clusters' (groups) of variables that are strongly inter-related

– the rotation is done to try to make variables **within a subgroup** score as highly (positively or negatively) as possible on one particular factor while

– ensuring that the loadings for these variables on the remaining factors are as low as possible.

In other words, the object of the rotation is to try to ensure that all variables have **high loadings only on one factor**.

2. Factor Rotation:

Goal: to improve the overall interpretability.

If there are 'clusters' (groups) of variables that are strongly inter-related

– the rotation is done to try to make variables **within a subgroup** score as highly (positively or negatively) as possible on one particular factor while

– ensuring that the loadings for these variables on the remaining factors are as low as possible.

In other words, the object of the rotation is to try to ensure that all variables have **high loadings only on one factor**.

Types of rotation method:

– orthogonal rotation the rotated factors will remain uncorrelated (varimax rotation)

– in oblique rotation the resulting factors will be correlated.

3. Calculation of factor scores:

When calculating the final factor scores (the values of the m factors, $F_1, F_2, \ldots, F_m$, for each observation), a decision needs to be made as to how many factors to include.

3. Calculation of factor scores:

When calculating the final factor scores (the values of the m factors, $F_1, F_2, \ldots, F_m$, for each observation), a decision needs to be made as to how many factors to include.

This is usually done using one of the following methods:

- Choose m such that the factors account for a particular percentage (e.g. 75%) of the total variability in the original variables.

3. Calculation of factor scores:

When calculating the final factor scores (the values of the m factors, $F_1, F_2, \dots, F_m$, for each observation), a decision needs to be made as to how many factors to include.

This is usually done using one of the following methods:

- Choose m such that the factors account for a particular percentage (e.g. 75%) of the total variability in the original variables.

- Choose m to be equal to the number of eigenvalues over 1 (if using the correlation matrix). [A different criteria must be used if using the covariance matrix.]

3. Calculation of factor scores:

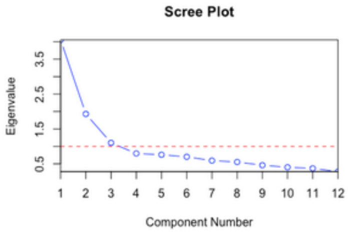
When calculating the final factor scores (the values of the m factors, $F_1, F_2, \ldots, F_m$, for each observation), a decision needs to be made as to how many factors to include.

This is usually done using one of the following methods:

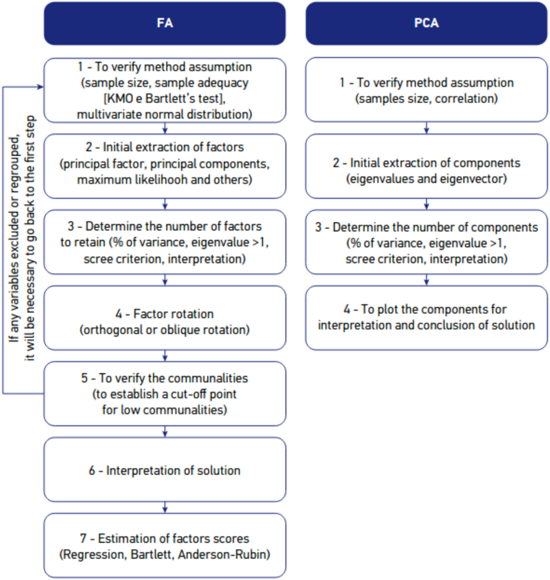
- Choose m such that the factors account for a particular percentage (e.g. 75%) of the total variability in the original variables.

- Choose m to be equal to the number of eigenvalues over 1 (if using the correlation matrix). [A different criteria must be used if using the covariance matrix.]

- Use the scree plot of the eigenvalues. This will indicate whether there is an obvious cut-off between large and small eigenvalues.



CFA – Common Factor Analysis



Summary

The difference between PCA and Factor Analysis

- PCA does not estimate specific effects, so it simply finds the mathematical definition of the "best" components (components who maximize variance).

- Factor Analysis will also estimate the components, common interpretable factors, and all of the specific factors.

Summary

The difference between PCA and Factor Analysis

- PCA does not estimate specific effects, so it simply finds the mathematical definition of the "best" components (components who maximize variance).

- Factor Analysis will also estimate the components, common interpretable factors, and all of the specific factors.

Different Goal

– PCA has as a goal to define new variables based on the highest variance explained and so forth.

– FA has as a goal to define new variables that we can understand and interpret in a business / practical manner.

Summary

The difference between PCA and Factor Analysis

- PCA does not estimate specific effects, so it simply finds the mathematical definition of the "best" components (components who maximize variance).

- Factor Analysis will also estimate the components, common interpretable factors, and all of the specific factors.

Different Goal

– PCA has as a goal to define new variables based on the highest variance explained and so forth.

– FA has as a goal to define new variables that we can understand and interpret in a business / practical manner.

Different Mathematical Model

– Both methods use decomposition, they differ in the details.

– CFA has an additional rotation of the final solution, while PCA does not.

Summary

The difference between PCA and Factor Analysis

- PCA does not estimate specific effects, so it simply finds the mathematical definition of the "best" components (components who maximize variance).

- Factor Analysis will also estimate the components, common interpretable factors, and all of the specific factors.

Different Goal

– PCA has as a goal to define new variables based on the highest variance explained and so forth.

– FA has as a goal to define new variables that we can understand and interpret in a business / practical manner.

Different Mathematical Model

– Both methods use decomposition, they differ in the details.

– CFA has an additional rotation of the final solution, while PCA does not.

Different Applications

– As Factor Analysis is more flexible for interpretation it is valuable in studies for marketing and psychology.

– PCA's allows for dimension reduction while keeping a maximum amount of information in a data set, hence, iused to simplify EDA or to preprocess data.