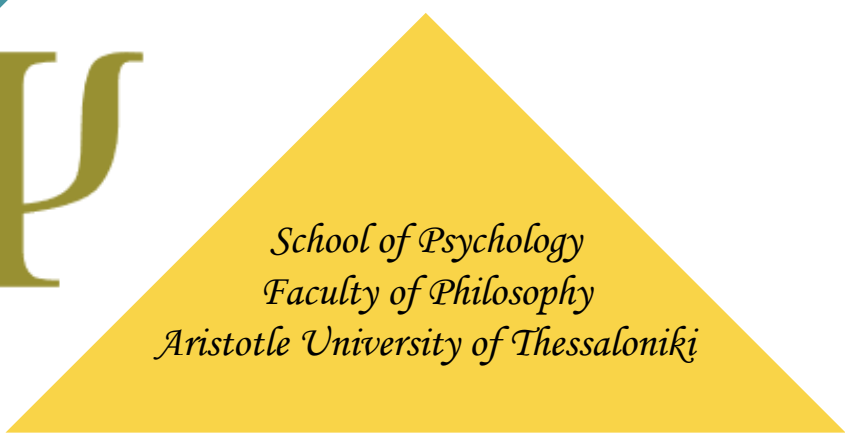





# Principal Component Analysis (PCA)

---

Konstantinos I. Bougioukas, MSc, PhD



*School of Psychology  
Faculty of Philosophy  
Aristotle University of Thessaloniki*

Principal Component Analysis (PCA) is a statistical technique used to **reduce the dimensionality of a dataset** while preserving as much of its original variability as possible.

PCA computes new variables, called **principal components**, which are obtained as **linear combinations** of the original variables.

$$PC1 : Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

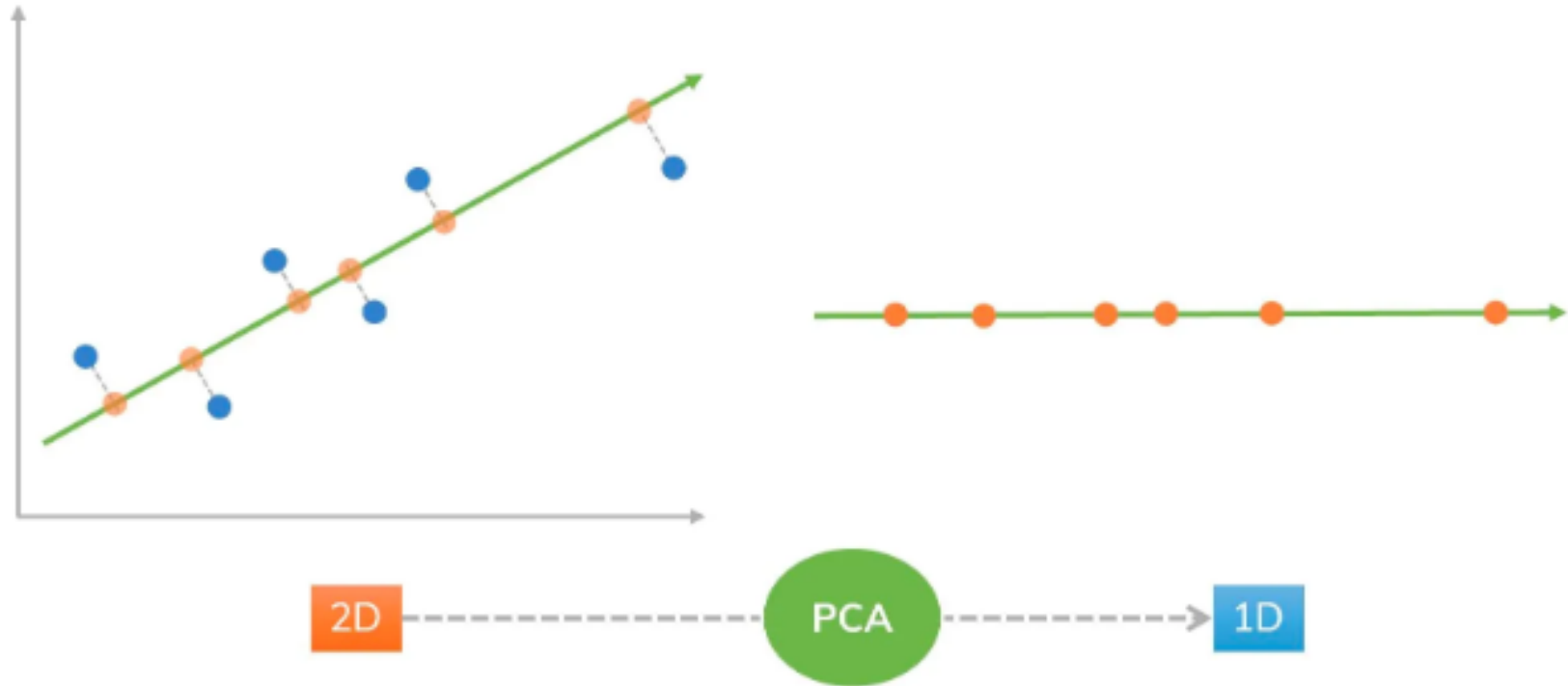
$$PC2 : Z_2 = \phi_{12}X_1 + \phi_{22}X_2 + \dots + \phi_{p2}X_p$$

...

...

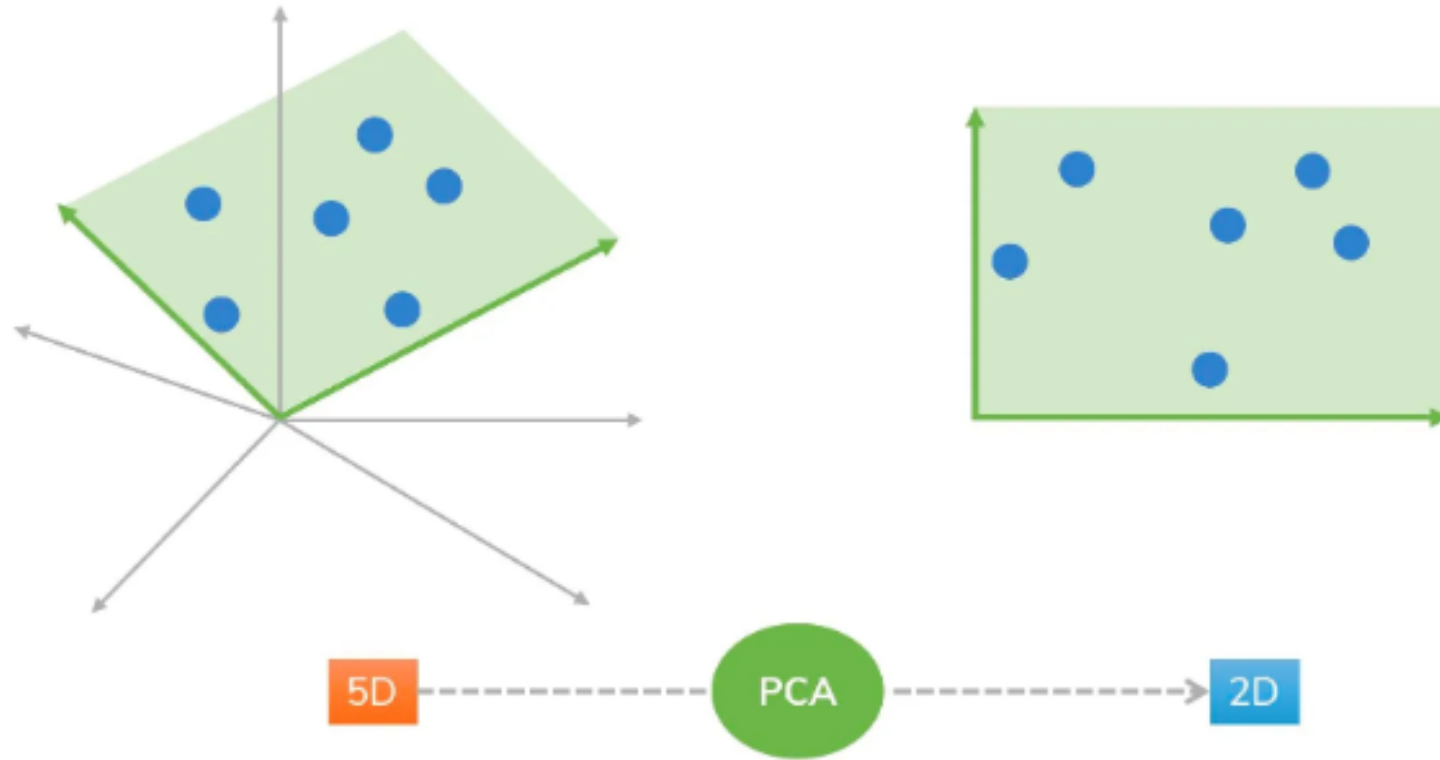
# From 2D to 1D

**Minimize the orthogonal distance to the line.**

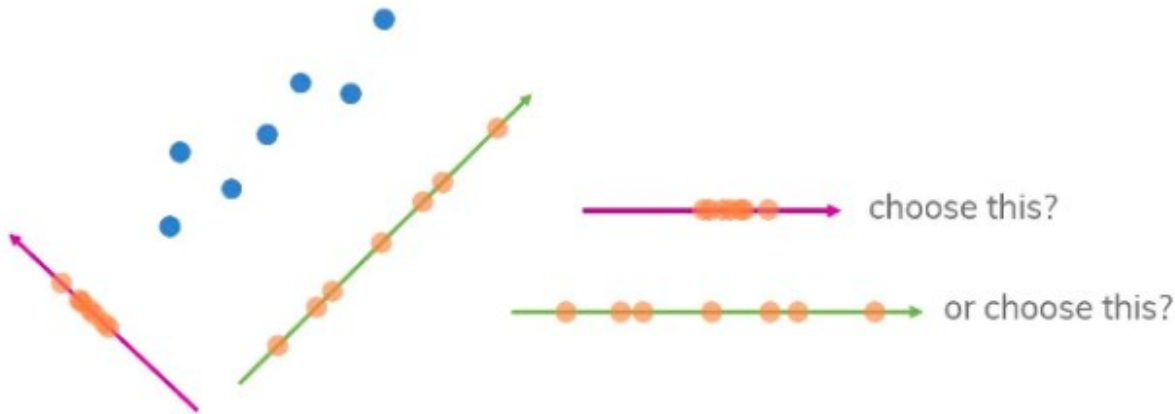


# From 5D to 2D

How can we choose the **green arrows** (their directions and their magnitudes)?



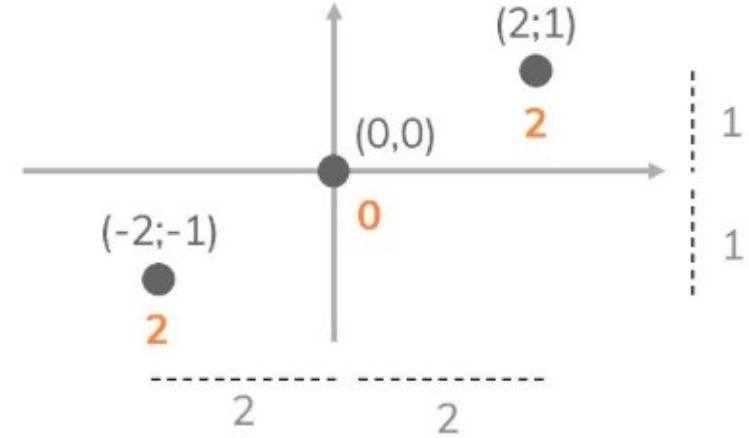
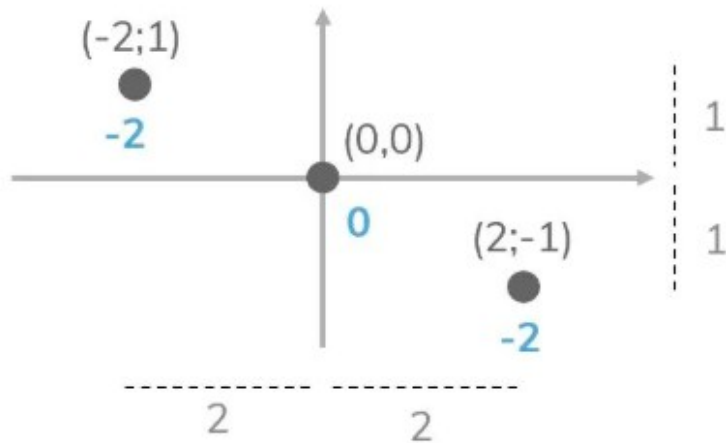
From a data points, there are many ways of projections, for example:



Maximize the Variance

Intuitively, the green line is better with more separated points. But how can we choose it "mathematically" (precisely)? We need to know about:

- **Mean:** finds the most balanced point in the data.
- **Variance:** measures the spread of data from the mean. However, variance is not enough. There are many different ways in that we get the same variance.
- **Covariance:** indicates the direction in that data are spreading.



Mean = 0

$$\text{x-variance} = \frac{2^2 + 0^2 + 2^2}{3} = \frac{8}{3}$$

$$\text{y-variance} = \frac{1^2 + 0^2 + 1^2}{3} = \frac{2}{3}$$

$$\text{Covariance} = \frac{(-2) + 0 + (-2)}{3} = -\frac{4}{3}$$

Mean = 0

$$\text{x-variance} = \frac{2^2 + 0^2 + 2^2}{3} = \frac{8}{3}$$

$$\text{y-variance} = \frac{1^2 + 0^2 + 1^2}{3} = \frac{2}{3}$$

$$\text{Covariance} = \frac{2 + 0 + 2}{3} = \frac{4}{3}$$

# PCA algorithm

1. Standardize the data
2. From the original data (a lot of features  $x_1, x_2, \dots, x_p$ ), we construct a covariance matrix  $C$ .
3. Compute the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_p$  and correspondent eigenvectors  $v_1, v_2, \dots, v_p$  of that matrix.
4. Select the top  $k < p$  couples  $\lambda$  and  $v$  (the highest eigenvalues) and we get a reduced projection matrix  $W_k$  ( $k$  eigenvectors as its columns; loadings, if scaled by  $\sqrt{\lambda}$ ).
5. Projection original data points to the  $k$ -dimensional plane. This step creates new data points on a new dimensional space ( $k$ ).

[illegible]
$$X_{nxp}$$

## Standardize the data

$$X_{n\chi\rho}^c$$

## Covariance matrix

$$C_{pxp} = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}$$

eigen  
decomposition

## Pairs of eigenvectors and eigenvalues

$v_1$	$\lambda_1$
$v_2$	$\lambda_2$
...	...

big

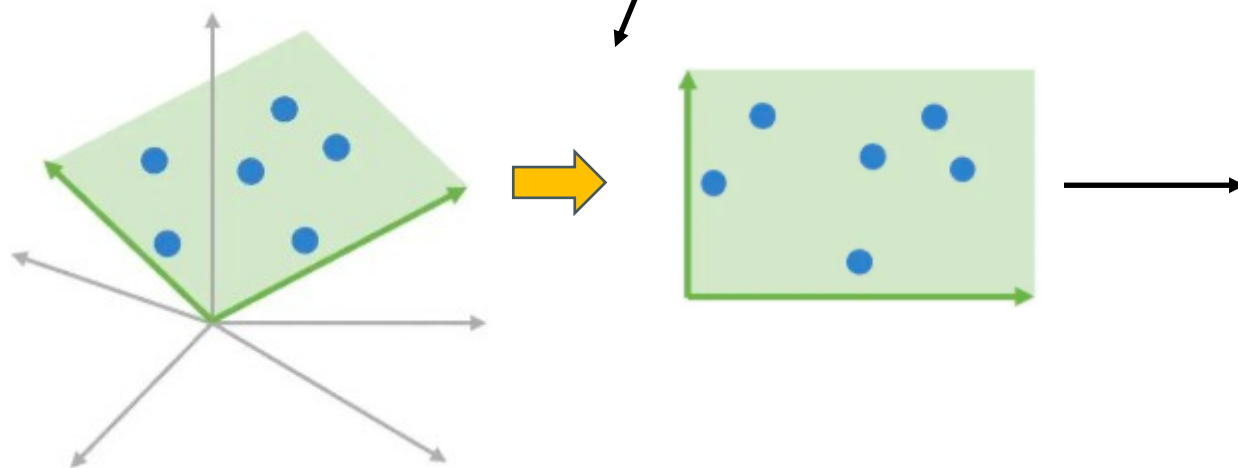
small

## Select the Top k Principal Components

$$W_{pxk}$$

## Component scores

### Small table

[illegible]



# Eigenvectors and Eigenvalues

**Eigenvectors matrix:**  $V_{p \times p}$

$$\mathbf{V} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_p \\ | & | & \dots & | \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1p} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pp} \end{bmatrix}$$

**Eigenvalue matrix:**  $\Lambda_{p \times p}$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}$$

**Loadings of first principal component: PC1**

$$\phi_{i1} = \alpha_{i1} \sqrt{\lambda_1} \quad i = 1, 2, \dots, p$$

$$\phi_{11}, \phi_{21}, \dots, \phi_{p1}$$

$$PC1 : Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

# Proportion of variance explained

To determine the proportion of total variance explained by the first principal component (i.e., direction by that component), we use the following formula:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_p}$$

# Component rotation

(not a part of classical PCA)

## Additional step

Rotation in PCA redistributes the variance between components, making the **loadings more distinct and interpretable**. In other words, rotation typically causes each observed variable to load highly on precisely one component. This clarifies the grouping of variables and makes the underlying dimensions easier to understand.

- **Orthogonal Rotation:** Methods like Varimax keep the components uncorrelated (i.e., orthogonal). This preserves the core mathematical properties of PCA, ensuring that the components remain independent.

(NOTE: the factor loadings can be interpreted as **correlation coefficients** between the original observed variables and the rotated components).

- **Oblique Rotation:** We don't use this type of rotation in PCA because the components are no longer orthogonal. Exploratory factor analysis (EFA) may be a more technically appropriate model.

# PC Loadings and variance

Component Loadings

	PC1	Component				Uniqueness
	1	2	3	4		
Obj1	0.574				0.632	
Obj2	0.581				0.613	
Obj3	0.531				0.679	
Obj4	0.604				0.596	
Obj5	0.532				0.635	
Obj6	0.489			0.428	0.576	
Obj7	0.517				0.675	
Obj8	0.572				0.531	
Obj9	0.468				0.732	
Obj10	0.545				0.674	
Marg1		0.784			0.367	
Marg2		0.615			0.499	
Marg3		0.596			0.602	
Marg4		0.559			0.614	
Marg5		0.601			0.575	
Marg6	0.368	0.403			0.671	
Marg7		0.511			0.714	
Str1			0.648		0.505	
Str2			0.381		0.724	

If a variable has high uniqueness (close to 1), it indicates that the variable is poorly explained by the selected number of components.

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	3.31	13.23	13.23
2	2.92	11.66	24.90
3	2.05	8.19	33.09
4	1.56	6.24	39.32

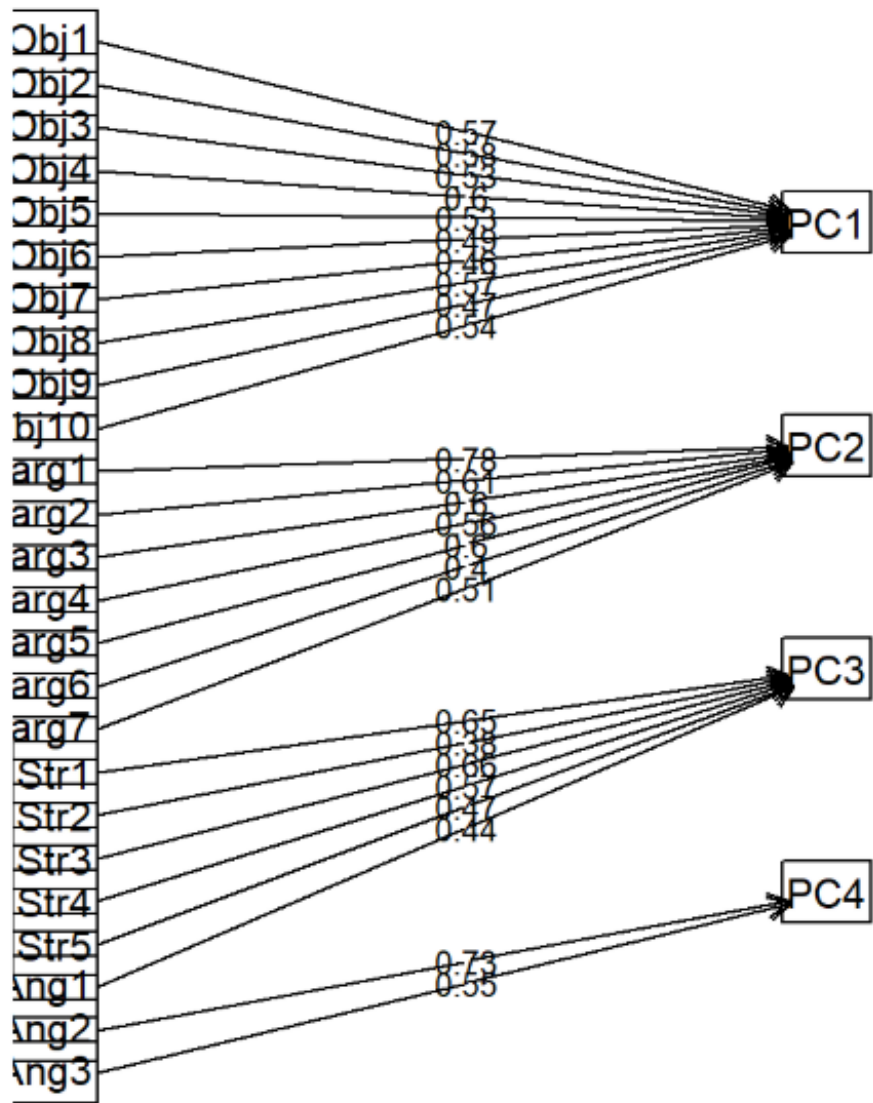
The SS Loadings column represents the **eigenvalues**  $\lambda$  for each component

Observed variables (features)

# Component Analysis diagram-Varimax rotation

Orthogonal rotation

Components Analysis



# PCA Assumptions

Identity matrix

$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- **Bartlett's test of sphericity**

A significant result ( $p < 0.05$ ) indicates that the correlation matrix significantly differs from an identity matrix. This suggests that the variables share enough correlation to justify the use of principal component analysis (PCA).

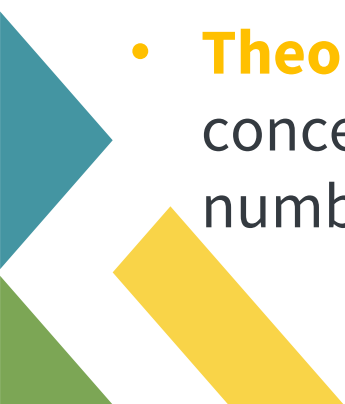
- **Kaiser-Meyer-Olkin (MKO) index of Sampling Adequacy**

(bare minimum of 0.5, values between 0.5 and 0.7 as mediocre, values between 0.7 and 0.8 as good, values between 0.8 and 0.9 as great, values above 0.9 are superb)

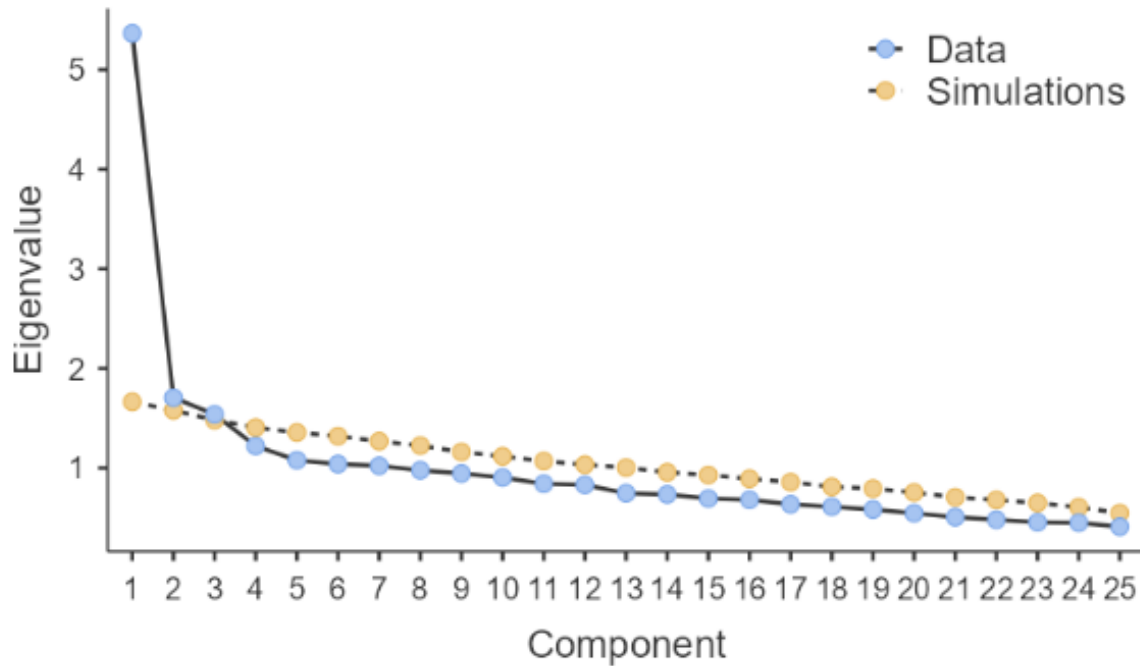
- **Address outliers.**

# How do we specify the Number of Components?

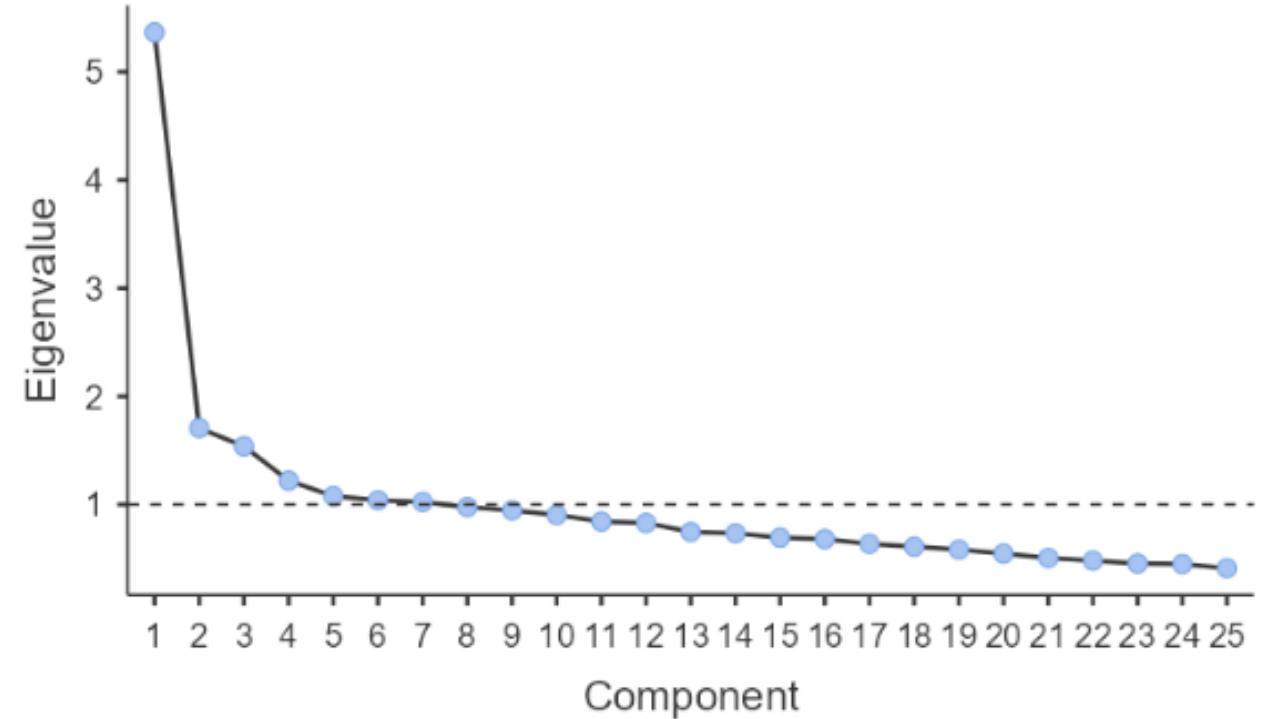
Our decisions on how many components to keep can be guided by several methods:

- **Cumulative variance explained** – Retain enough components to account for a desired proportion of total variance.
  - **Parallel analysis** – Compare the observed eigenvalues with those obtained from randomly generated data to determine which components are meaningful.
  - **Kaiser's criterion** – Retain components with **eigenvalues greater than 1**.
  - **Theoretical justification** – Retain components that align with prior knowledge, conceptual frameworks, or hypotheses relevant to the domain of study (fixed number).
- 

# Scree Plot



Scree plot determining the number of components based on parallel analysis.



Scree plot determining the number of components based on Kaiser's criterion.