PCA (*Principal Component Analysis*)

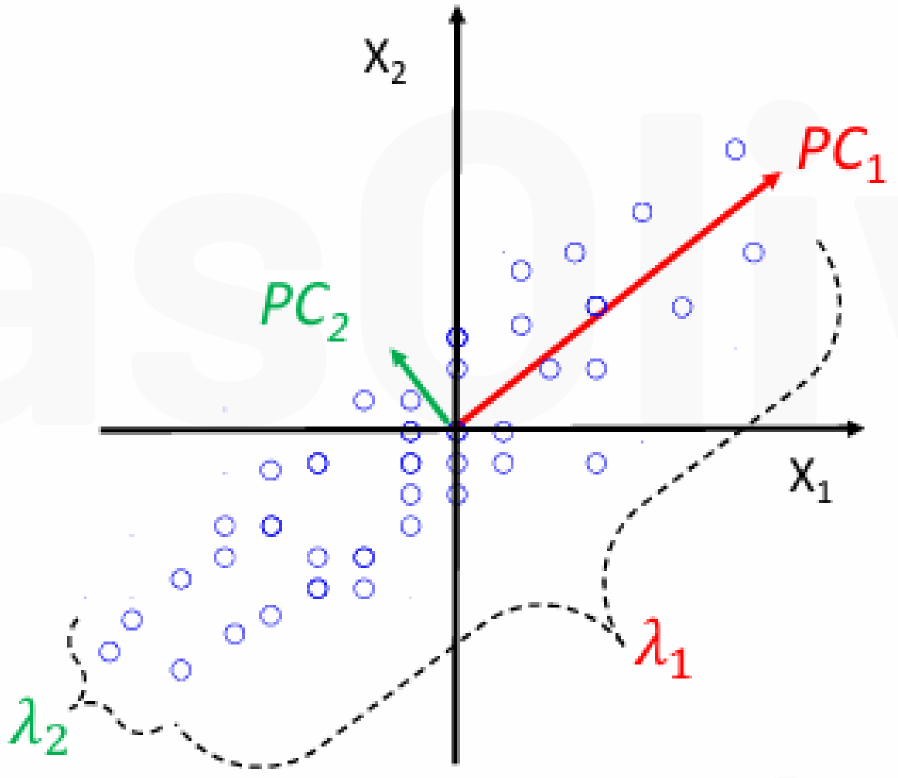
What is it?

A dimensionality reduction technique that:

Transforms the original variables $x_1,\dots,x_D$ into orthogonal linear combinations named <u>Principal Components</u> (PCs) defined by:

$$\boxed{u_j = a_{j1}x_1 + \cdots + a_{jD}x_D = \boldsymbol{a}_j^T\boldsymbol{x}}$$

$\boldsymbol{a}_j^T$ are component loadings—correlations (ranging from -1 to 1) between each original variable and the component.



Implying information loss!

Assumptions

- Linearity.
- Orthogonality.
- Second-order statistics, variance, is the most important.

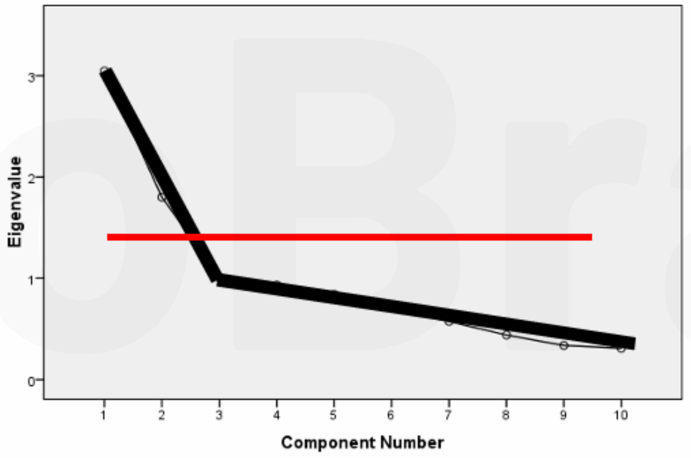
1 Should I use it?

- Observe sample size to # of variables ratio
  - From 5:1 to 30:1
- Scale/ standardize variables
- Compute the Correlation matrix
  - Correlation > 0.3
- Perform Bartett's test of sphericity
  - p-value <0.05, then PCA may be helpful
- Apply KMO (*Kaiser-Meyer-Olkin*) measure of sampling adequacy (MSA)
  - MSA > 0.6

2 How many PCs?

The fewer, the better!

- Kaiser criterion
  - Eigenvalues > 1
- Explained variance
  - At least 60%
- Scree Plot

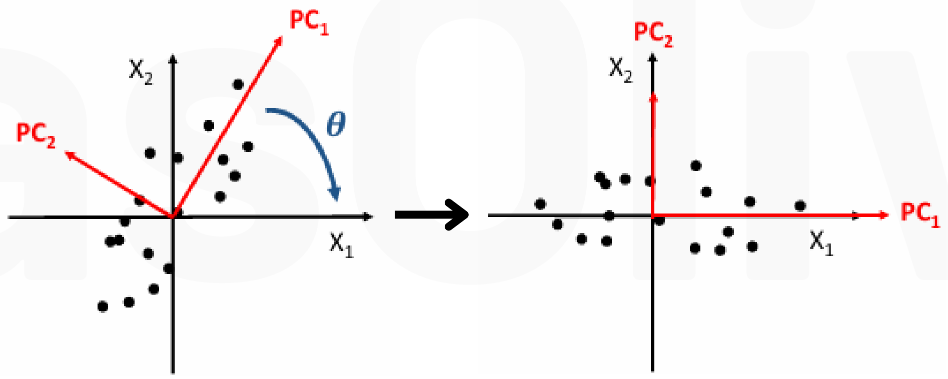


3 How to interpret?

Components are defined by the variables that load highly on them.

Rotations allow simpler solutions by linking each variable to fewer PCs.

Varimax, Oblique (Oblimin, Promax), Quartimax, Equamax.



4 Obtain component scores

Scores correspond to the new set of variables