BADJI MOKHTAR UNIVERSITY
FACULTY OF ECONOMICS AND MANAGEMENT SCIENCES
Department: Financial Sciences
Solutions of Serie3.
Data Analysis, Statistics, and Applications

Level: L3
Instructor: Dr. Soumaya Allaoui

Exercise 1:

- **What are the objectives of factor analysis methods?**

To reduce the dimensionality of data by identifying underlying factors or components that explain the observed variability in the data.

The main objective of PCA is to reduce the dimensionality of a quantitative dataset while preserving as much variability (informations) as possible. It transforms the original variables into a new set of uncorrelated variables called principal components.

- **What is the role of eigenvalues and eigenvectors in PCA?**

Eigenvalues indicate the amount of variance explained by each principal component. and Eigenvectors define the directions of the new principal components in the feature space (i.e., the weights used to combine original variables).

- **What are principal components, and how are they ordered?**

Principal components are new variables obtained as linear combinations of the original variables.

They are ordered by the amount of variance they explain:

1. The first principal component (PC1) captures the most variance.
2. The second (PC2) captures the next most, and so on.

- **What is the indicator of association between quantitative variables?**

In PCA, we work with quantitative variables.

If variables are measured on the ***same scale***, we use the ***covariance matrix*** to analyze the association.

If variables are on ***different scales or have large differences in their standard deviations***, we standardize them and use the ***correlation matrix*** instead, to ensure each variable contributes equally to the analysis.

- **When should we apply the operations of centering and scaling to the data before PCA?**
  - Centering (subtracting the mean) is always necessary.
  - Scaling (standardizing to unit variance) is recommended when variables are measured on different scales or have different variances. to ensure that no variable dominates due to its scale.

- **Explain the use of the Chi-square test for Correspondence Analysis (CA).**

The Chi-square test is used to assess the independence of rows and columns in a contingency table, helping to

interpret the associations between categorical variables in CA.

- Explain the main difference between the methods: Principal Component Analysis (PCA), Correspondence Analysis (CA), and Multiple Correspondence Analysis (MCA).

The difference on **the type of the given datasets**

I. Principal Component Analysis (PCA)

a) ***Data type:*** *Used for quantitative (continuous) variables.*

b) ***Association indicators:***

i. Use the covariance matrix when variables are measured on the same scale.

ii. Use the correlation matrix when variables have different scales or large differences in standard deviations.

c) ***Objective:*** *Reduce dimensionality of Quantitative variables by finding linear combinations (principal components) that explain the maximum variance.*

d) ***Output:*** *Eigenvalues (variance explained) and eigenvectors (directions of components).*

II. Correspondence Analysis (CA)

a) ***Data type:*** Used for ***two qualitatives variables*** (cross-tabulated in a contingency table).

b) ***Association indicator:*** The chi-squared distance is used to measure the association between rows and columns.

c) ***Objective:*** *Display the associations between qualitatives variables in a low-dimensional space.*

d) ***Output:*** *Coordinates of rows and columns that best represent associations.*

III. Multiple Correspondence Analysis (MCA)

Data type: Used for ***more than two qualitatives variables*** (extended CA for multiple variables).