

Multivariate Statistical Modeling

Dimensionality Reduction, Factor Analysis and Data Matrix Architecture

Giorgio Saibene

Level: Undergraduate / Bachelor of Science

Course Overview

This foundational course introduces the mathematical mechanics of corporate and research data matrix architectures. Students will master core exploratory multivariate techniques, focusing on identifying latent structures, eliminating collinearity variables, and reducing data dimensionality without systemic information loss.

Core Methodological Modules

1. Data Matrix Foundations & Preprocessing

Geometric representation of multivariate observations in n -dimensional spaces. Missing data imputation methods, detection of multivariate outliers via Mahalanobis distance, and validation of assumptions: normality, homoscedasticity, and linearity.

2. Principal Component Analysis (PCA)

Mathematical derivation of variance maximization. Eigenvalue and eigenvector decomposition of covariance and correlation matrices. Component rotation paradigms (Varimax vs. Oblimin) and optimal factor score extraction techniques.

3. Exploratory Factor Analysis (EFA)

The latent variable construct paradigm. Distinctive mathematical differences between PCA and Common Factor Analysis. Evaluation of sampling adequacy via Kaiser-Meyer-Olkin (KMO) metrics and Bartlett's test of sphericity.

4. Canonical Correlation Analysis (CCA)

Simultaneous modeling of multiple dependent and independent metric variables. Derivation of

canonical functions, canonical loadings, and cross-loadings. Statistical significance testing of canonical roots.

5. Cluster Analysis & Segmentation Baselines

Hierarchical clustering algorithms (Ward's method) and non-hierarchical optimization (K-means). Mathematical distance measures (Euclidean, Manhattan). Validation and profiling of cluster stability across distinct sample subgroups.