Applied Multivariate Analysis – Syllabus

September 2025 – December 2025

Instructor: An-Shun Tai

Institute of Statistics and Data Science, National Tsing Hua University

（戴安順, 清大統計與數據科學）

Course Outline

Part I. Foundations

Lecture 1. Introduction and Review (Chapters 1–4)

- Nature and scope of multivariate analysis; applications across disciplines
- Review of vectors, matrices, and quadratic forms relevant for multivariate methods
- Properties of the multivariate normal distribution (marginals, conditionals, linear combinations)

Lecture 2. Inference about a Mean Vector (Chapter 5)

- Sampling distributions: sample mean vector and covariance matrix; Wishart distribution
- Hotelling' s T^2 statistic for inference on mean vectors
- Likelihood ratio tests for hypotheses on mean vectors and covariance structures

Part II. Dimension Reduction

Lecture 3. Principal Component Analysis (PCA) (Chapter 8)

- Eigenvalues and eigenvectors of the covariance (or correlation) matrix
- Principal components as linear combinations that maximize variance
- Interpretation via proportion of variance explained, scree plots, and biplots
- Applications in reducing dimensionality while retaining structure

Lecture 4. Factor Analysis (Chapter 9)

- The common factor model: decomposing observed variation into common and unique parts
- Distinction between PCA and FA in terms of model assumptions
- Factor loadings and communalities
- Rotation methods (orthogonal and oblique) for interpretability
- Estimation and goodness-of-fit diagnostics

Lecture 5. Multidimensional Scaling (MDS) and Nonlinear Embedding (Chapter 12.6 + Supplement)

- Classical (metric) MDS based on Euclidean distances
- Non-metric MDS using monotone transformations of dissimilarities
- t-SNE: stochastic neighbor embedding to preserve local structure (supplement)
- UMAP: graph-based manifold learning with improved scalability (supplement)
- Comparison of linear vs nonlinear dimension-reduction techniques

Part III. Classification and Clustering

Lecture 6. Discriminant Analysis (Chapter 11)

- Fisher's linear discriminant functions for two and multiple groups
- Quadratic discriminant analysis when group covariance matrices differ
- Assessment of classification rules: apparent error rates, cross-validation

Lecture 7. Classification Extensions (Chapter 11 + Supplement)

- Logistic regression for binary, multinomial, and ordinal outcomes
- Overview of modern classifiers (decision trees, support vector machines) (supplement)
- Evaluation metrics: misclassification rate, ROC curves, AUC, confusion matrices

Lecture 8. Cluster Analysis (Chapter 12)

- Hierarchical clustering: distance metrics, linkage methods, dendrogram interpretation
- Non-hierarchical clustering: k-means, k-medoids, and validation via silhouette widths
- Model-based clustering: Gaussian mixture models and EM algorithm

Part IV. Regression and Canonical Methods

Lecture 9. Multivariate Regression (Chapters 6 and 7)

- Model specification with multiple dependent variables
- Estimation of regression coefficients and covariance of errors
- Inference via Wilks' Lambda, Pillai's trace, Hotelling–Lawley trace, and Roy's largest root
- Relationship between multivariate regression and MANOVA

Lecture 10. Canonical Correlation Analysis (CCA) (Chapter 10)

- Canonical variates: linear combinations of X and Y that maximize correlation
- Extraction of successive canonical correlations subject to orthogonality constraints

- Testing the significance of canonical correlations with likelihood ratio tests
- Interpretation through canonical loadings and redundancy analysis

Lecture 11. Correspondence Analysis (Chapter 12.7 + Supplement)

- Simple correspondence analysis for two-way contingency tables
- Multiple correspondence analysis for categorical survey data (supplement)
- Geometric representation of row and column profiles
- Relationship of CA to PCA on categorical data

Part V. Advanced and Modern Topics

Lecture 12. Structural Equation Models (SEM) (Supplement)

- Path diagrams linking observed and latent variables
- Specification, identification, and estimation of SEMs
- Examples from psychology and social sciences

Lecture 13. Partial Least Squares (PLS) (Supplement)

- Motivation: prediction in the presence of multicollinearity and many predictors
- Latent components chosen to maximize covariance between predictors and responses
- Comparison of PCA, CCA, and PLS in terms of objectives and applications
- Applications in chemometrics and bioinformatics

Prerequisites

Linear algebra, calculus, probability, statistical inference, and linear models. Some experience with R programming is desirable, but not strictly required.

Textbook

Johnson, R. A., & Wichern, D. W. (2018). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson.

Reference Books

Rencher, A. C., & Christensen, W. F. (2012). *Methods of Multivariate Analysis* (3rd ed.). Wiley.
Everitt, B. S., & Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer.

Grading

Homework Assignments (30%; approximately 4–5 assignments), Midterm Examination (30%), and Final Project (40%).

Final Project Guidelines

The final project will be conducted as a group presentation in the last week of the semester. In principle, each group should consist of five students. The presentation must be **delivered in English** and should include the following components:

1. Description of the dataset
2. Motivation of the research problem
3. Choice of methods and methodological introduction
4. Analysis results
5. Conclusion

On the presentation day, one specific part of the report will be randomly assigned to each group for presentation.

Teaching Assistants

- 卓辰樺　Email: hichcho@gapp.nthu.edu.tw
- 朱宸儀　Email: