Lecture 15

Review

DSA 8070 Multivariate Analysis
November 29 - December 3, 2021

Whitney Huang
Clemson University

Objectives of Multivariate Analysis

Dimensionality Reduction or Structural Simplification

- Goal: to reduce the "dimensionality" by considering a small number of (linear) combinations of a large number of measurements without losing important information

- Techniques:

  - **Principal Component Analysis**

  - **Factor Analysis**

  - **Multidimensional Scaling**

Grouping or Classification

- Goal: to **identify** groups of "similar" units or to **classify** units into previously defined groups
- Techniques:
  - **Classification Analysis**
  - **Cluster Analysis**

Investigation of the Dependence among Variables and Prediction

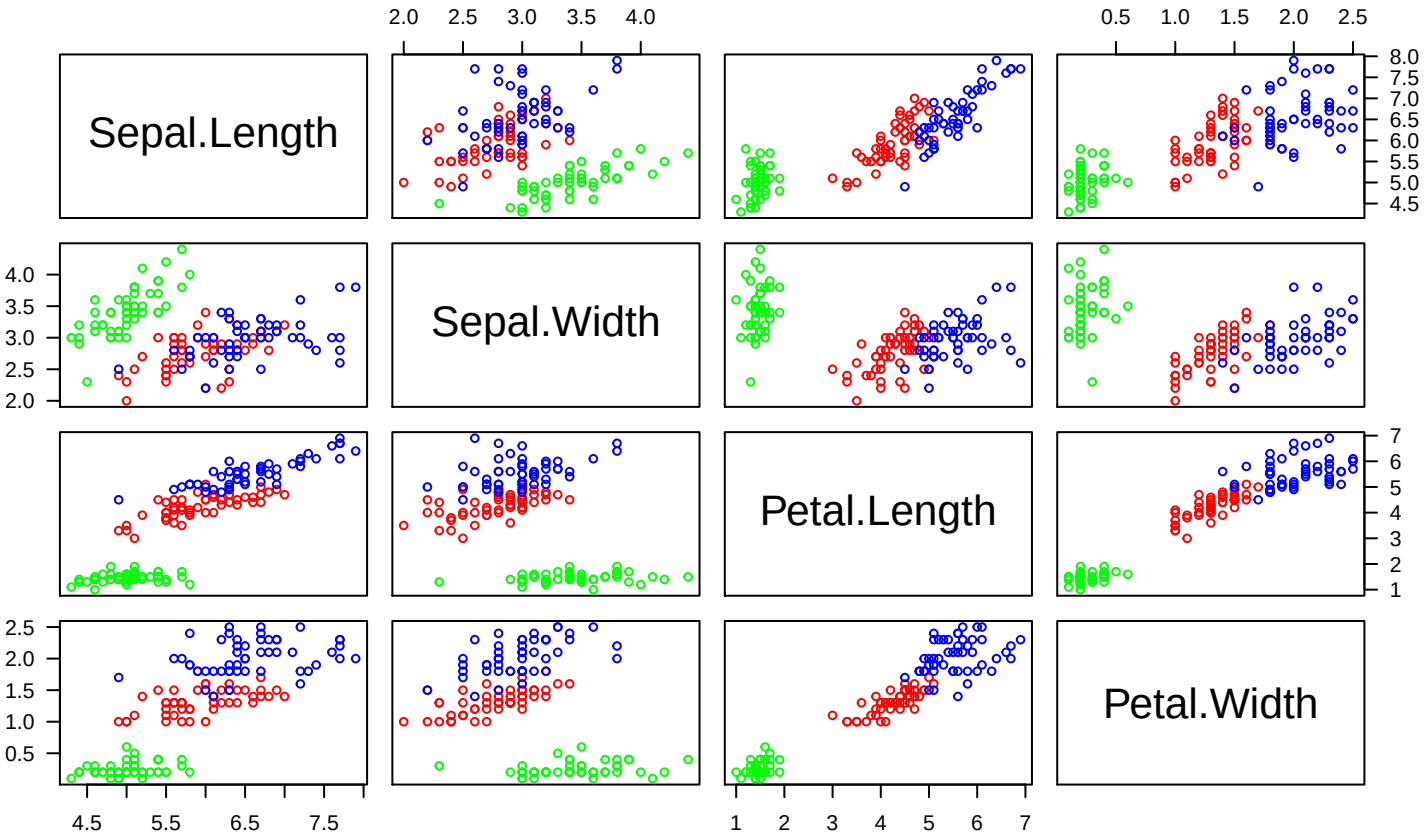
- Goal: to estimate the relationship among variables and, if the variables are associated, to predict the value of some of them given information on the others
- Techniques:
  - **Multivariate Regression**
  - **Canonical Correlation Analysis**

Hypothesis Testing

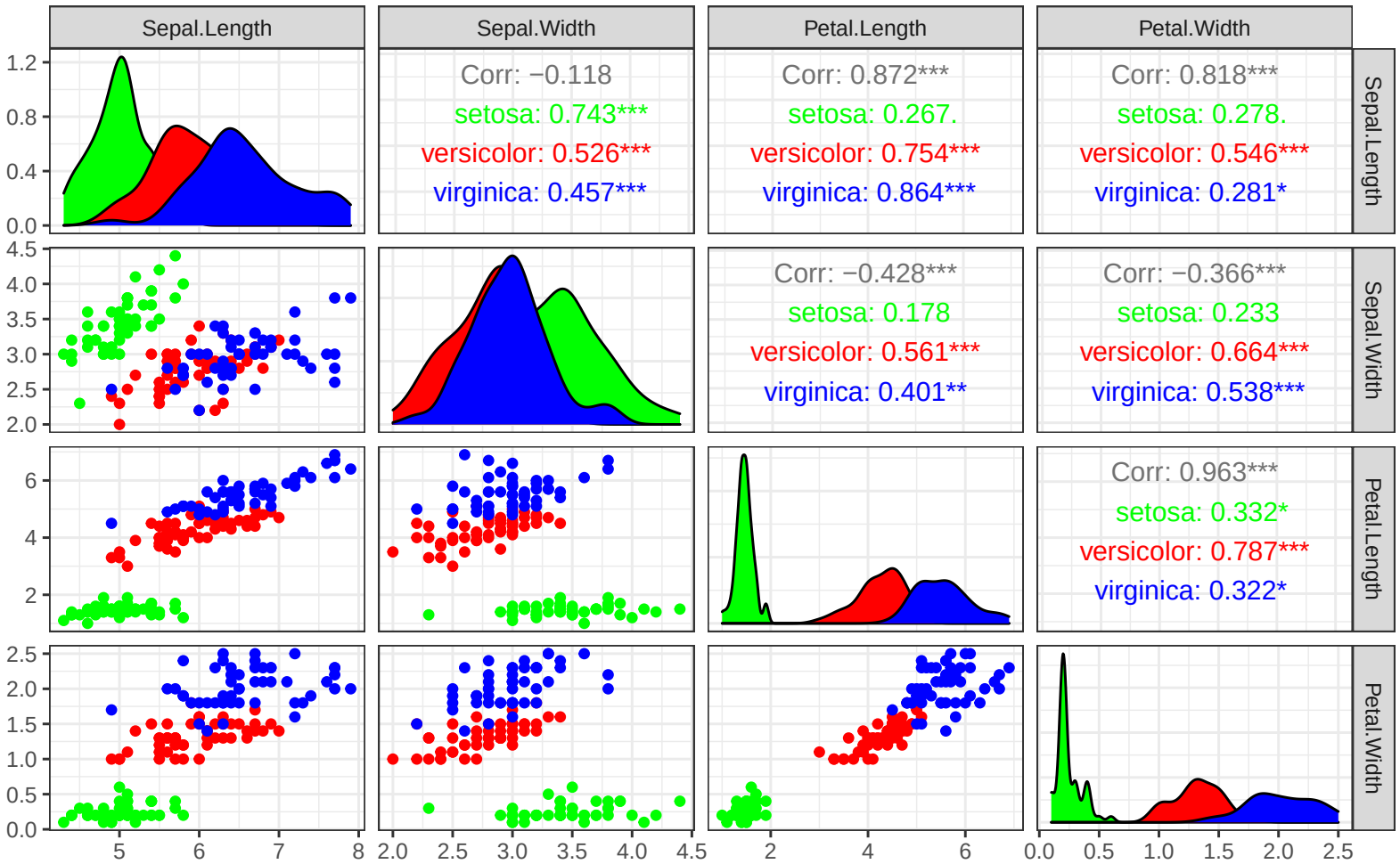
- Goal: to test if differences in sets of response mean vectors for two or more groups large enough to be distinguished from sampling variation
- Techniques:
  - **Hotelling's T^2**
  - **MAVONA**

Data Visualization

Scatterplot Matrix



Scatterplot Matrix using *ggpairs*



Summary of Multivariate Procedures

Summary of Multivariate Procedures (adapted from C.J.Anderson UIUC Soc 584 & Psych 594)

	PCA	CCA	MANOVA	(Linear) DA
Data	$p \geq 2$ variables	2 sets of $p \geq 2$ and $q \geq 2$ variables	$p \geq 2$ variables with groups/factors	$p \geq 2$ variables with group variable
Requirements	None	$\boldsymbol{S}$ are positive definite	$\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}_k, \Sigma)$	Equal Σ_k
Focus	Within set of variables	Between sets of variables	Between groups relative to within groups	Between groups relative to within groups
Goals	Account for as much variances as possible	Determine nature and strength of the relationship between sets	Statistical inferences regarding $\boldsymbol{\mu}_k$	Classification

Similarities Between Multivariate Procedures (adapted from C.J.Anderson UIUC Soc 584 & Psych 594)

- All seek linear combinations of the original variables that maximize some criterion
- All techniques use the inter-relationship between variables (i.e., covariance or correlation matrix)
- All try to reduce the dimensionality of the problem and thus aid in the description and interpretation of relations between variables
- Geometrically, all methods can be thought of as finding (or studying) sub-spaces of the original higher dimensional space

Topics Not Addressed

- Inference of Covariance Structure
- Multivariate Analysis of Categorical Data (Ref. [Bishop, Fienberg, Holland, 1975])
- Support-Vector Machine (Ref. Ch. 9 of [James, Witten, Hastie, and Tibshirani, 2021])
- Nonlinear Dimensionality Reduction and Manifold Learning (Ref. Ch. 16 of [Izenman, 2008])

Additional Readings

- An Introduction to Multivariate Statistical Analysis by T. W. Anderson
- Modern Multivariate Statistical Techniques, 2008 by A. J. Izenman
- An Introduction to Statistical Learning: with Applications in R, 2021 (2nd ed.) by G James, D Witten, T Hastie, and R Tibshirani
- Principal Component Analysis, 2002 by I. T. Jolliffe