

# KHOA CÔNG NGHỆ THÔNG TIN-ĐHKHTN

## PHÂN TÍCH THỐNG KÊ DỮ LIỆU NHIỀU BIẾN

Giảng viên: PGS.TS. Lý Quốc Ngọc  
TPHCM, 8-2020



KHOA CÔNG NGHỆ THÔNG TIN  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

**fit@hcmus**

# PHÂN TÍCH THỐNG KÊ DỮ LIỆU NHIỀU BIẾN

## Bài giảng 1: Giới thiệu về PTTKDLNB

Giảng viên: PGS.TS. Lý Quốc Ngọc



KHOA CÔNG NGHỆ THÔNG TIN  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

[fit@hcmus](mailto:fit@hcmus)

# Nội dung

- **Chương 1.** Tổng quan về PTTKDLNB
- **Chương 2.** Các khái niệm cơ bản trong PTTKDLNB
- **Chương 3.** Mô hình hồi qui tuyến tính nhiều biến.
- **Chương 4.** Phân tích thành phần chính
- **Chương 5.** Phân tích dữ kiện
- **Chương 6.** Phân tích tương quan chính tắc
- **Chương 7.** Gom cụm
- **Chương 8.** Phân lớp

# Nội dung

## Chương 1. Tổng quan về PTTKDLNB

**1.1.** Mục đích của phân tích dữ liệu nhiều biến

**1.2.** Các vấn đề chính trong phân tích dữ liệu nhiều biến

**1.3.** Ứng dụng của phân tích dữ liệu nhiều biến

# Nội dung

## Chương 2. Các khái niệm cơ bản trong PTTKDLNB

2.1. Đại lượng ngẫu nhiên

2.2. Vectơ trung bình và ma trận phương sai

2.3. Hàm mật độ phân bố chuẩn nhiều biến

2.4. Một số định luật cơ bản

# Nội dung

## Chương 3. Mô hình hồi qui tuyến tính nhiều biến

**3.1.** Mô hình hồi qui tuyến tính truyền thống

**3.2.** Mô hình hồi qui tuyến tính nhiều biến

# Nội dung

## Chương 4. Phân tích thành phần chính

### 4.1. Mục đích

### 4.2. Tìm các thành phần chính của dữ liệu nhiều biến

### 4.3. Xấp xỉ dữ liệu nhiều biến bởi các thành phần chính

### 4.4. Khảo sát tính chất các thành phần chính

## Chương 5. Phân tích dữ kiện

### 5.1. Mục đích

### 5.2. Mô hình dữ kiện

### 5.3. Ước lượng thông số của mô hình dữ kiện



# Nội dung

## Chương 6. Phân tích tương quan chính tắc

### 6.1. Mục đích

### 6.2. Biến chính tắc và tương quan chính tắc

### 6.3. Phân tích tương quan chính tắc

# Nội dung

## Chương 7. **Gom cụm**

**7.1.** Phương pháp gom cụm phân cấp

**7.2.** Phương pháp gom cụm không phân cấp

# Nội dung

## Chương 8. **Phân lớp**

**8.1.** Phân lớp dựa trên phân bố chuẩn

**8.2.** Ước lượng hàm phân lớp

**8.3.** Phương pháp Fisher

# Chương 1

## Tổng quan về PTTKDLNB

**1.1. Mục đích của phân tích dữ liệu nhiều biến**

**1.2. Các vấn đề chính trong phân tích dữ liệu nhiều biến**

**1.3. Ứng dụng của phân tích dữ liệu nhiều biến**

## 1.1. Mục đích của phân tích dữ liệu nhiều biến

### - Đối tượng sử dụng

Các nhà nghiên cứu trong các lĩnh vực khoa học thường xuyên xử lý dữ liệu nhiều biến (Biological, Physical, Social Sciences, Computer Science, vv.)

- **Trình bày các khái niệm và phương pháp** phân tích dữ liệu nhiều biến dựa trên xác suất thống kê và đại số ma trận, giải tích.

- **Khảo sát khoa học là quá trình học lặp** (iterative learning process), đạt được bởi tiếp nhận và phân tích dữ liệu.

## 1.1. Mục đích của phân tích dữ liệu nhiều biến

- Giảm chiều dữ liệu (Data **reduction**)
- Sắp xếp và nhóm dữ liệu (Sorting and **Grouping**)
- Khảo sát sự phụ thuộc giữa các biến (Investigation of the **dependence** among variable).
- Tiên đoán (**Prediction**)
- Xây dựng giả thiết và kiểm tra giả thiết (Hypothesis **construction and testing**)

## 1.2. Các vấn đề chính

- **Hồi qui tuyến tính nhiều biến.**  
(Multivariate linear regression models)
- **Phân tích thành phần chính.**  
(Principal components analysis)
- **Phân tích dữ kiện.**  
(Factor Analysis)
- **Phân tích tương quan chính tắc.**  
(Canonical correlation analysis)
- **Gom cụm** (Clustering)
- **Phân lớp** (Classification)