

## データの分析

### 1 データの代表値

- (1) データの整理
- (2) 平均値
- (3) 中央値・最頻値

### 2 データの散らばり

- (1) 四分位数
- (2) 箱ひげ図

### 3 分散と標準偏差

- (1) 基本
- (2) 平均値と分散
- (3) 変数（変量）変換

### 4 データの相関

- (1) 散布図・相関表
- (2) 相関係数

# 1 データの整理と分析

## (1) データの整理

- (i) 階級 → データを整理するために用いる区間
  - ～の幅 → 区間の幅
  - ～値 → 階級の真ん中の値
  
- (ii) 度数 → それぞれの階級に入っているデータの値の個数
  - ～分布 → 各階級に度数を対応させたもの
  - ～表 → それを表にしたもの → それをグラフにしたのが「ヒストグラム」
  - 相対～ → 各階級の度数を度数の合計で割った値  
(相対度数) = (その階級の度数) / (度数の合計)

## (2) 代表値

- (i) 平均値  $\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n$ 
  - (ア) 度数分布から調べるとき  $\bar{x} = (x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n) / n$
  - (イ) 仮平均  $x_0$  を用いるとき  $\bar{x} = \{(x_1 - x_0) + (x_2 - x_0) + \dots + (x_n - x_0)\} / n$
  
- (ii) 中央値 (メジアン) → すべてのデータの値を小さい順に並べたとき、中央の順位にくる値  
(データの個数が  $2n$  個のときは、 $n$  番と  $(n+1)$  番の平均値)
  
- (iii) 最頻値 (モード) → データを度数分布表に整理したとき、度数が最も大きい階級の階級値