



## 慶應義塾大学ビジネス・スクール

# 統計学 ノート (1)

## 1 はじめに

統計学はもともと社会における現象の法則性に対する人間のさまざまな実践的関心から起こった。現象の法則性を知るために、すべてを詳しく調べ、規則性から法則を見つけてもよいし、また、一部を観察して、そこから論理に基づいた推論を行い、全体の法則性を発見してもよい。

全体の法則性を知るために、すべてを調べなければならないという考え方は全数調査と呼ばれ、実際には国勢調査などで用いられている。

一方、現在の統計学は「一部」から「全体」を知ろうとする、科学的推論のための方法論の理論的体系となっている。この方法論は、確率論に基づく数学的根拠とともに、広い応用範囲と妥当性をもっている。

ここで、統計の目的は得られたデータからデータを集めたもとの集団についての情報を引き出すことである。このとき、得られたデータ(観測値)の集まりを標本(sample)といい、もととなる集まりを母集団(population)という。

この標本と母集団を区別して考え、標本から母集団への橋渡しとして確率論が用いられるようになって、近代統計学の基礎が確立した。そして、これに基づいて、統計的推論の論理が築かれた。母集団に対する推定(estimation)の理論と仮説検定(hypothesis testing)である。

## 2 標本

### 2.1 データの収集、分類とその記述

母集団に関する妥当な結論を得るためには、どのようにして母集団から標本を抽出すればよいかということが問題である。その標本抽出の方法は、母集団を構成するどの個体も標本に選ばれる確率が同じになるような抽出法であり、無作為抽出という。

データは量的データ、質的データに分けられる。数値などで定量的に表されるデータを量的データとよぶ。また、量的データは連続型変数と離散型変数に分けられる。長さ、重さ、温度、時間などのある区間内の任意の値をとりうる変数は連続型変数であり、事故の件数、人数、金額などは離散型変数の例である。数値として観測することができず、ある