

第1章 統計の基礎概念

1.1 母集団と標本

20歳代の日本人男性の全体を考えよう。これは、有限個の個体からなる人間の集まりであるが、これを、その世代の身長（という特性値）を知りたいという視点から調査の対象とすると、あるいは、その世代の政治意識を知る目的で、支持政党あり（特性値0）なし（特性値1）の調査対象とすると、その集団は観測値の源泉としての集団とみることができる。このような視点からみた集団を母集団という。この問題を解くには、母集団をすべて調べてみればわかる。これを全数調査という。これは原理的には可能だが、実際には調査コストや時間などを考えると実行は難しい。たとえば、工場ですべての製品を次々と大量に生産される製品の不良品の割合を知りたい場合、母集団はこれらの製品全体で事実上無限母集団と考えられ、全数調査は不可能である。

このように我々が知りたいのは、母集団に関する情報であるが、その母集団をしらみつぶしに調べることはできないという立場にある。したがって、母集団の一部だけを調べて母集団に関する結論を引き出す方法を考えざるを得ない。母集団のいくつかの個体を選び特性値を観測することを標本をとるといふ。それが、 n 個の観測値からなるとき、大きさ n の（1つの）標本が得られたという。統計とは母集団に関する結論を標本から引き出す方法である。母集団において、各要素（個体）の特性値はそれぞれ違っている。したがって、我々が知りたいのは、特性値の母集団における分布（母集団分布）、あるいはその分布を特徴づけるいくつかの量ということになる。

さてこのとき標本のとり方が問題となる。我々の目的は、得られた標本に基づき特性値の母集団分布を知ることであるから、どのような方法で標本をとってもよいというわけにはいかない。たとえば、ある特定の政党の集会上で述べたような政治意識調査を行えば、値0の割合が実際より多くなり、よい標本でないことは明らかであろう。すなわち、その標本から得られる分布は、母集団分布の縮図と見なすことができない。このようなことを避けるには、どの男性も標本として選ばれるチャンス（確率）を同じにすることで、このような標本の選び方を無作為標本抽出法といい、得られた標本を無作為標本という。