

第1節

測定値の特徴を捕える  
「記述統計」

■ 本章の流れ

本章では図1の流れに沿って「確率」および「統計」の基礎について解説します。この分野の内容は多岐に渡りますが、本章では統計学全般で重要な役割を果たす「正規分布」を理解することを目標とします。

● 記述統計と推測統計

統計学をおおまかに分類すると、図2のようになります。手元にあるデータを効率よく整理する方法について扱う分野を「記述統計」(descriptive statistics)といいます。これは、データの特徴を「平均」や「分散」といった1つの数値で代表させ、把握しやすくする技法です。手元に大量のデータ(いわゆるビッグ・データ)があるときに威力を発揮します。

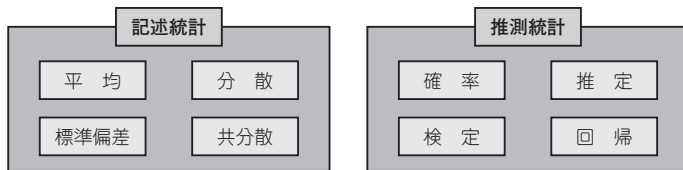
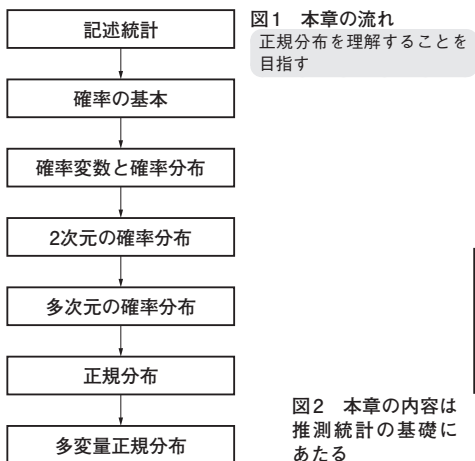


図2 本章の内容は推測統計の基礎にあたる

(a) 手元のデータを整理する

(b) 手元のデータをもとにして、未取得のデータを推測する

表1 ゲーム1の度数分布表

スコア	人数
1点	1人
2点	3人
3点	1人
4点	3人
5点	4人
6点	6人
7点	5人
8点	3人
9点	3人
10点	1人

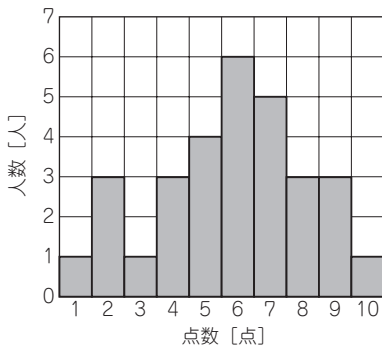


図3 ゲーム1の結果のヒストグラム

表2 ゲーム2の度数分布表

スコア	人数
1点	6人
2点	4人
3点	2人
4点	1人
5点	2人
6点	2人
7点	2人
8点	3人
9点	5人
10点	3人

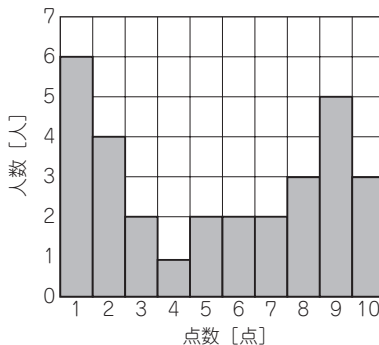


図4 ゲーム2の結果のヒストグラム

これに対して、実際の研究・調査では少数のデータ(スモール・データ)しか手に入らないことがほとんどです。その「一部」のデータから、「全体」の挙動を推測する手法として発展したのが「推測統計」(inferential statistics)です。推測統計では、推定結果の信頼度を表現するために「確率」の考え方が使われます。

本章では、確率論の初歩および正規分布の扱い方について絞って解説します。これは、推測統計の基礎という位置づけになります。

■ 記述統計

● ヒストグラム

まずは、入手したデータの整理法について確認します。ここでは例として、点数が1点から10点までの「ゲーム1」を、30人がやったとします。このゲーム1の結果にもとづいて各スコアに対応する人数を集計し、表1に示す「度数分布表」(frequency distribution table)が得られたとします。また、この度数分布表を柱状グラフで表すと、図3の「ヒストグラム」(histogram)になります。

また、別の「ゲーム2」(表2)を30人がやって、表2に示す度数分布表と図4に示すヒストグラムが得られたとします。

● 平均

入手した多数のデータをそのまま把握するのは困難なので、統計学ではいろいろな「代表値」(representative value)が使われます。代表値は、取