1.2　母集団分布

特性値が離散量なら，その母集団分布は棒グラフで表示される．たとえば，20 歳代日本人男性の無党派層（値 1）の比率が 0.43 であるすると，それは図 1.1 のようになる．もちろん，我々は，この比率が 0.43 であることは知らない．そこで，母集団から大きさ n の標本[1] $X_1, X_2, \ldots, X_n$ を無作為に抽出することにしよう．このとき，各 X_i が 値 1 である可能性（確率）はどれも同じ 0.43 であり，しかも，それらは互いに影響しあうことはない（独立）と考えられる．この標本に基づき，母集団分布についての統計的推測を行うのである．たとえば，実際に $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ というデータが得られたとき，各 x_i の値は 0 または 1 であるから，無党派層の比率は，データの平均

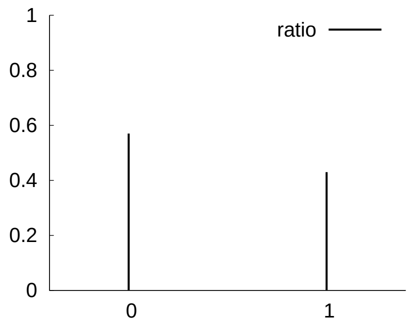

図 1.1: 無党派率の母集団分布

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\#\{x_i : x_i = 1\}$$

（ここで，$\#\{x_i : x_i = 1\}$ は値が 1 であったデーターの**度数**）で推定できそうであるということは，直観的に理解できるであろう．しかし，この方法で得られる値が信頼できるものであるかかどうかは検証の必要がある．これも統計学のテーマの一つである．

特性値が身長のように連続量の場合は，母集団分布は，たとえば身長を 2 cm 間隔で階級に分けるとすれば，図 1.2 のような柱状のグラフ，**ヒストグラム**で表示される．ここで，縦軸の目盛りは単位長さあたりの相対度数を表している．たとえば，区間 $[168, 170]$ に対応する長方形の面積は $2 \times 0.06 = 0.12$ で，その階級に 12% の男性が属していることを示している．した

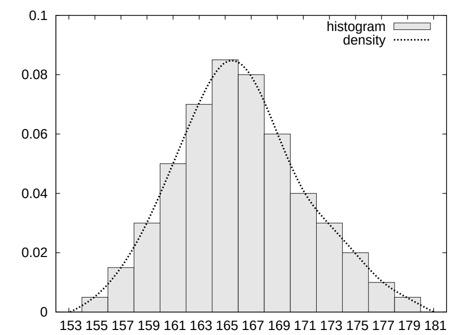

図 1.2: 身長の母集団分布ヒストグラム

[1] 標本は，実際に観測するまではどのような値が現れるかわからない．すなわち，“変数” である．これを，明示的に大文字で表し，実際に出現した値（データ，実測値，観測値などという）を対応する小文字や具体的な数値で表すことが多い．