

§ 1.1 度数分布

ある範囲の事物の各々の属性の程度を表す数量を変量という。例えば、日本人 19 歳男子は数多くいて、各人に背丈という属性がある。そこで、日本人 19 歳男子について、背丈という属性の程度を表す数量つまり身長は、変量と考えることができる。

ある範囲のものの各々の属性の程度表す変量について、その範囲のものの幾つかについてその変量を測定・観測などして得られた数値の集まりをデータ (data) という¹⁾。例えば、変量として日本人 19 歳男子の身長を考えると、日本人 19 歳男子の何人かの身長を実際に測定して得られた数値の集まりがデータである。

データの要素が実数で表される量であるとする。

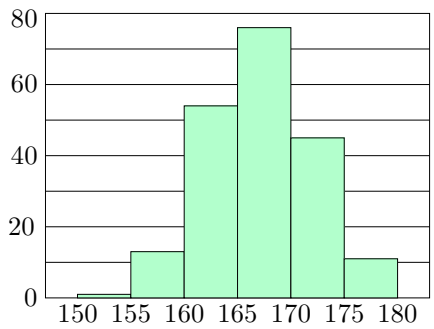
変量としてある高専の 4 年生の身長を例にとる。身長の単位は cm とする。例えば 4 年生 200 人の身長を測れば、要素が 200 個のデータを得ることができる。しかし、データの 200 個の要素を単に並べられただけでは、要素の分布の状況が分かりにくい。このようなとき、データの要素の分布の状況を分かり易くするために、度数分布表とかヒストグラムなどが用いられる。

度数分布表とは、例えば次のように、データの要素の範囲を区間で分割して、各区間に属す要素の個数を表にしたものである。

身長の区間（単位は cm）（階級）	人数（度数）	人数の割合（相対度数）
150 以上 155 未満	1	0.5%
155 以上 160 未満	13	6.5%
160 以上 165 未満	54	27%
165 以上 170 未満	76	38%
170 以上 175 未満	45	22.5%
175 以上 180 未満	11	5.5%
計	200	100%

表の左側の欄にある要素の区間を階級または単に級という。階級の幅を階級幅または級間隔といい、階級の中央の値を階級値という。例えば、165 以上 170 未満の階級では、階級幅は 5 で、階級値は $\frac{165+170}{2} = 167.5$ である。また、表の真中の欄にある、各階級に属す要素の個数を度数 (frequency) という。更に、表の右側の欄にある、各階級に属す要素の個数の割合を相対度数 (relative frequency) という。

更に度数分布表からできる右図のようないくつかの長方形をくっつけた形のグラフを、ヒストグラム (histogram) あるいは柱状グラフをという。ヒストグラムでは、通常、横軸を階級に分割して縦軸に各階級に対する度数あるいは相対度数をとる。



¹⁾ 数値でないものを含むデータもあるがここでは数値化されるものだけを考える。