

データの分析

いくつかのデータを、基準をもとに区間ごとに分ける。それを表に整理したものを**度数分布表**といい、棒グラフに整理したものを**ヒストグラム**という。

このとき、データを分析するための用語がいくつかある！

区間のことを**階級**、区間の真ん中の値を**階級値**、区間のデータの個数を**度数**という。

$$\text{相対度数} = \frac{\text{各階級の度数}}{\text{全ての度数の合計}}$$

はじめの階級からその階級までの度数の合計を**累積度数**という。

データの特徴を表す数値を**代表値**という。
平均値、中央値、最頻値、範囲などがある。

平均値 … (データの数値の合計) ÷ (データの個数)

中央値 … データを大小順に並べたときに、真ん中のデータの値

最頻値 … データの中で、一番多い数値

範囲 = (データの最大値) - (データの最小値)

例えば、20 人の身長を、10cm ごとの区間に分けてみる。

階級	階級値	度数	相対度数	累積度数
140 ～ 150	145	5	$5 \div 20 = 0.25$	5
150 ～ 160	155	10	$10 \div 20 = 0.50$	15
160 ～ 170	165	5	$5 \div 20 = 0.25$	20

平均値 : $(145 \times 5 + 155 \times 10 + 165 \times 5) \div 20 = 155$

中央値 : 真ん中は 10 人目なので、155

最頻値：度数 10 が一番大きいので、155

ちなみに、階級値は、階級の両端の値をたして、2 でわると求まる。

例えば、140～150 という階級ならば、階級値は $\frac{140 + 150}{2} = 145$ となる。