1 ▶ 相対度数分布グラフ

1-1 資料をグラフに整理しよう
—— 度数分布表、ヒストグラム

相対度数分布グラフって、どう描くの？

これから資料の整理の仕方と資料の特徴を捉える数値の導き方を紹介しましょう。

ある中学校のクラスでテストをしました。20 人のテストの結果が次のようだったとします。この資料を整理してみましょう。

2人　3人　5人

43、47、52、52、54、61、67、67、68、69、

6人　3人　1人

70、71、71、73、76、78、82、84、84、91

何十点台が何人いるか調べてみます。

40 点台は 2 人います。点数が 40 点台であれば、点数が 40 点以上 50 点未満として表に整理しましょう。

階級	階級値	度数
40 以上　50 未満	45	2
50 以上　60 未満	55	3
60 以上　70 未満	65	5
70 以上　80 未満	75	6
80 以上　90 未満	85	3
90 以上 100 以下	95	1

ここで、この例を参考にして、統計学の用語を紹介します。

● **サイズ（大きさ）**

この場合は、調査対象の人数、資料に含まれる個数、20 人のことです。本書では、わかりやすく資料の個数といったりします。

● **変量**

この場合は、テストの点数。資料の各々の値です。

● **階級（class interval）**

40 点以上 50 点未満といった区切りのこと。

● **階級値（midpoint）**

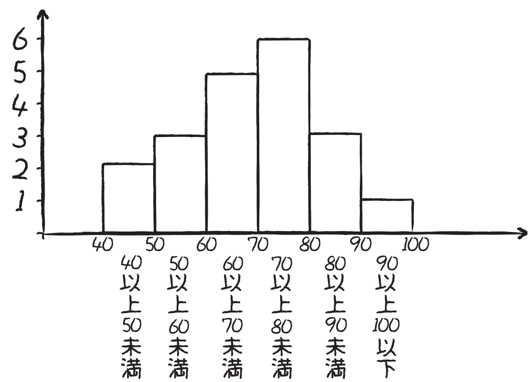
40 点以上 50 点未満の階級であれば、40 点と 50 点の平均の 45 点です。midpoint という英語からもわかるように階級の範囲の真ん中の値です。

● **度数（frequency）**

この場合、階級ごとの人数のこと。

資料を階級ごとに整理して度数を調べたものを**度数分布表（frequency table）**といいます。

この表をもとに各階級の度数を柱状グラフで表すと次のようになります。



このように、度数分布表を柱状グラフにして表現したものを**ヒストグラム（histogram）**といいます。

ヒストグラムを描くと資料の特徴がひと目でわかり便利です。ただ、このままでは資料同士を比較するときに不便です。人数の異なる A 中学校と B 中学校の得点分布を比べたいときは、階級に含まれる度数の全体に対する割合を比較した方がよいでしょう。

資料全体に対する度数の割合を**相対度数（relative frequency）**といいます。

度数分布表をもとにして、相対度数を表に書き込むと次のようになります。このような表を相対度数分布表といいます。

$\frac{2}{20}$

階級	階級値	度数	相対度数	累積相対度数
40 以上 50 未満	45	2	0.1	0.1
50 以上 60 未満	55	3	0.15	0.25
60 以上 70 未満	65	5	0.25	0.5
70 以上 80 未満	75	6	0.3	0.8
80 以上 90 未満	85	3	0.15	0.95
90 以上 100 以下	95	1	0.05	1
		20	1	

相対度数は、全体の資料数を 1 としたときの階級の割合を表します。資料のサイズ（資料の個数）が 20 で、40 点以上 50 点未満の階級の度数は 2 であることから、40 点以上 50 点未満の階級の相対度数は、

$$\frac{2}{20} = 0.1$$

となります。

これに対して、相対度数を足しあげたものを累積相対度数といいます。60 点以上 70 点未満の累積相対度数は、それより低い階級の相対度数を足しあげて、

$$0.1 + 0.15 + 0.25 = 0.5$$

と計算します。

累積相対度数の最後の欄は 1 になっていますね。

これは偶然ではありません。相対度数をすべて足しあげると、つねに 1 になるからです。相対度数とは、全体の資料数を 1 としたときの割合なのですから、すべての相対度数を足すと全体の割合 1 になるわけです。

相対度数、累積相対度数のことを抽象的な書き方でまとめておきます。

階級	階級値	度数	相対度数	累積相対度数
第 1 階級	x_1	f_1	$\frac{f_1}{N}$	$\frac{f_1}{N}$
第 2 階級	x_2	f_2	$\frac{f_2}{N}$	$\frac{f_1+f_2}{N}$
第 3 階級	x_3	f_3	$\frac{f_3}{N}$	$\frac{f_1+f_2+f_3}{N}$
⋮	⋮	⋮	⋮	⋮
第 n 階級	x_n	f_n	$\frac{f_n}{N}$	1
		N	1	