【中学数学】 第５章 データの活用

体系数学 2/数研出版/代数編

Riku Sugawara

12.2025

第 1 節　データの整理

1.1　度数分布表とヒストグラム

Definition 1.1.1　度数分布表

データの範囲を適当に区切ったとき, 各区間に含まれるデータの個数を **度数** といい,
各区間にその区間の度数を対応させて整理した右のような表を **度数分布表** という.

階級（cm）	度数（人）
135 以上〜140 未満	2
140〜145	4
145〜150	5
150〜155	8
155〜160	11
160〜165	9
165〜170	7
170〜175	4
計	**50**

表 1.1.1　身長の度数分布表

Definition 1.1.2　階級

度数分布表において, 区切られた各区間を **階級**, 区間の幅を **階級の幅**,
各階級の中央の値を **階級値** という.

Definition 1.1.3　ヒストグラム

度数分布表を, 柱状のグラフで表したものを **ヒストグラム** という.
ヒストグラムの各長方形の横の長さは階級の幅を表し, 高さは各階級の度数を表す.

Definition 1.1.4　度数折れ線

ヒストグラムの各長方形の上の辺の中点を結んでできる折れ線グラフを **度数折れ線** という.
ただし, 度数折れ線をつくる時は, ヒストグラムの左右の両端に度数が 0 の階級があるものと考える.

例題 【教 p.114 追加例題 1】

表 1.1.1 において, 次の問いに答えなさい.

(1) 階級の幅は何か.

(2) 階級の個数はいくつか.

(3) 階級 140cm 以上 145cm 未満の度数はいくつか.

(4) 階級 140cm 以上 145cm 未満の階級値はいくつか.

1.2 相対度数

Definition 1.2.1 相対度数

度数の合計に対する各階級の度数の割合を **相対度数** という. 相対度数はふつう小数を用いて表す.

Theorem 1.2.1 相対度数

相対度数は次の式で求められる.

$$\text{(相対度数)} = \frac{\text{(その階級の度数)}}{\text{(度数の合計)}}$$

Definition1.2.1 より, 相対度数の合計は必ず **1** になる.

また, 相対度数を使用するメリットは, 度数の合計が異なる複数の分布について, **より正確に比較することができる** ことである.

1.3 累積度数

Definition 1.3.1 累積度数

度数分布表において, 各階級以下または各階級以上の階級の度数をたし合わせたものを**累積度数** という.

また, 累積度数を表にまとめたものを **累積度数分布表** という.

Definition 1.3.2 累積相対度数

累積度数についても, 度数の合計に対する各階級の累積度数の割合を考えることがある.

この割合を **累積相対度数** という.

例題 【教 p.119 追加例題 2】

表 1.1.1 において, 次の問いに答えなさい.

（1）各階級の相対度数を求めなさい.

（2）累積度数の列を加え, 累積度数分布表を作成しなさい.

（3）累積相対度数を求めなさい.

（4）身長が 150cm 以上の人の割合は何 % か求めなさい.

Proof.

階級（cm）	**度数（人）**	**相対度数**	**累積度数 (人)**	**累積相対度数**
135 以上〜140 未満	2	0.04	2	0.04
140〜145	4	0.08	6	0.12
145〜150	5	0.10	11	0.22
150〜155	8	0.16	19	0.38
155〜160	11	0.22	30	0.60
160〜165	9	0.18	39	0.78
165〜170	7	0.14	46	0.92
170〜175	4	0.08	50	1.00
計	**50**	1.00		

表 1.3.1　身長の累積度数分布表

□

第 2 節　データの代表値

Definition 2.0.1　代表値

いくつかの値が集まったデータがあるとき, そのデータ全体の特徴を表す数値を, そのデータの **代表値** という.

代表値には, 平均値, 中央値, 最頻値がある.

2.1　平均値

Definition 2.1.1　平均値

n 個の値が集まったデータがあるとする.

これら n 個の値の合計を個数 n でわった値を, このデータの **平均値** という.

Theorem 2.1.1　平均値

平均値は次の式で求められる.

$$\text{(平均値)} = \frac{\text{(データの値の合計)}}{\text{(データの個数)}}$$

練習問題 【教 p.120 練習 5】

次のデータは, ジョギングを日課にしている A さんが最近 5 日間に行ったジョギングの時間である. このデータの平均値を求めよ.

23　　18　　35　　27　　42　　(単位は 分)

Proof.

$$\begin{aligned}\text{(平均値)} &= \frac{23+18+35+27+42}{5} \\ &= \frac{145}{5} \\ &= 29(\text{分})\end{aligned}$$

□

Definition 2.1.2　中央値

データを大きさの順に並べたとき, その中央にくる値を, **中央値** または **メジアン** という.
ただし, データの個数が偶数のとき, 中央に 2 つの値が並ぶから, その 2 つの値の平均値を中央値とする.

練習問題 【教 p.120 練習 6】

次のデータは, あるクラスの生徒 10 人の英語のテストの得点である. このデータの中央値を求めよ.

75　38　49　88　61　83　44　67　58　95　(単位は 点)

Proof.
10 人の英語のテストの得点を, 低い順に並べると

38　44　49　58　61　67　75　83　88　95

5 番目と 6 番目の得点の平均値が中央値であるから

$$
\begin{aligned}
\textbf{(中央値)} &= \frac{61+67}{2} \\
&= 64(\text{点})
\end{aligned}
$$

□

Definition 2.1.3　最頻値

データにおいて, 最も個数の多い値を, そのデータの **最頻値** または **モード** という.
データが度数分布表に整理されているときは, **度数が最も大きい階級の階級値を最頻値**とする.

練習問題 【教 p.121 練習 7 改訂】

表 1.1.1 において, 身長の最頻値を求めよ. このデータの最頻値を求めよ.

度数分布表を利用したデータの平均値を求める方法を考えよう．データが度数分布表にまとめられているとき，**ある階級に含まれるデータは，すべてその階級の階級値をとるものと考えて** 平均値を求める．

Theorem 2.1.2　度数分布表を利用した平均値

度数分布表を利用した平均値は次の式で求められる．

$$(\text{平均値}) = \frac{\{(\text{階級値}) \times (\text{度数})\}\text{ の合計}}{(\text{度数の合計})}$$

例題 【教 p.121 追加例題 3】

表 1.1.1 において，次の問いに答えなさい．

(1) 身長の平均値を求めなさい．

(2) 身長の中央値を求めなさい．

Proof.

階級（cm）	度数（人）	階級値	階級値 × 度数
135 以上〜140 未満	2	137.5	275
140〜145	4	142.5	570
145〜150	5	147.5	737.5
150〜155	8	152.5	1220
155〜160	11	157.5	1732.5
160〜165	9	162.5	1462.5
165〜170	7	167.5	1172.5
170〜175	4	172.5	690
計	**50**		7860

$$(\text{平均値}) = \frac{7860}{50} = 157.2(\text{cm})$$
$$(\text{中央値}) = 157.5(\text{cm})$$

□

第 3 節　データの散らばりと四分位範囲

3.1　範囲

Definition 3.1.1　範囲

データのとる値のうち, 最大のものから最小のものをひいた値を **範囲** または **レンジ** という.

範囲は, データの散らばりの程度を表す.

Theorem 3.1.1　範囲

範囲は次の式で求められる.

$$(\text{範囲}) = (\text{最大値}) - (\text{最小値})$$

3.2　四分位数

次のデータは, ある中学校の A 組と B 組のそれぞれ 20 人に行ったテストの結果を, 得点の低い順に並べたものである. (単位は点)

A 組のテスト結果

26	31	35	39	44	48	52	55	57	63	67	68	74	75	78	82	85	87	92	93

B 組のテスト結果

27	42	47	49	51	53	54	56	59	64	66	68	69	69	72	76	79	82	85	94

表 3.2.1

Definition 3.2.1　四分位数

データの値を大きさの順に並べたとき, 4 等分する位置にくる値を **四分位数** という. 四分位数は小さい方から順に **第 1 四分位数**, **第 2 四分位数**, **第 3 四分位数** という. 第 2 四分位数は中央値のことである.

例題 【教 p.124 例 2】

表 3.2.1 において, 四分位数を求めよ.

Proof.

中央値すなわち第 2 四分位数は

$$\frac{63+67}{2}=65(\text{点})$$

第 1 四分位数は

$$\frac{44+48}{2}=46(\text{点})$$

第 3 四分位数は

$$\frac{78+82}{2}=80(\text{点})$$

□

練習問題 【教 p.124 練習 10】

表 3.2.1 において, B 組のデータの四分位数を求めよ.

Proof.

中央値すなわち第 2 四分位数は

$$\frac{64+66}{2}=65(\text{点})$$

第 1 四分位数は

$$\frac{51+53}{2}=52(\text{点})$$

第 3 四分位数は

$$\frac{72+76}{2}=74(\text{点})$$

□

3.3 四分位範囲

Definition 3.3.1 四分位範囲

第 3 四分位数から第 1 四分位数をひいた差を **四分位範囲** という.

四分位範囲は, 範囲よりもより 中央値に近いところ でのデータの散らばりの程度を調べることができる.

Theorem 3.3.1 四分位範囲

四分位範囲は次の式で求められる.

$$\textbf{(四分位範囲)} = \textbf{(第 3 四分位数)} - \textbf{(第 1 四分位数)}$$

第 1 四分位数と第 3 四分位数の間の区間には, データ全体のほぼ半分 が入っており, データの中には極端に大きな値や小さな値があっても, 影響を受けにくい.

一般に, データが中央付近に集中しているほど, 四分位範囲は小さくなり, データの散らばりの程度は小さいといえる.

例題 【教 p.125 例 3】

表 3.2.1 において, 四分位範囲を求めよ.

Proof.
第 1 四分位数は 46 点, 第 3 四分位数は 80 点であるから,

$$\textbf{(四分位範囲)} = 80 - 46 = 34(\text{点})$$

□

練習問題 【教 p.125 練習 11 改訂】

表 3.2.1 において, 次の問いに答えよ.

(1) B 組のデータ四分位範囲を求めよ.

(2) A 組と B 組の四分位範囲から, データの散らばりの程度が大きいのはどちらの組みであると考えられるか. 理由も含めて答えよ.

Proof.
(1)B 組のデータの四分位範囲は

$$\textbf{(四分位範囲)} = 74 - 52 = 22(\text{点})$$

(2)A 組の四分位範囲は 34 点, B 組の四分位範囲は 22 点であるから,A 組のデータの四分位範囲の方が大きい.
よって, データの散らばりの程度が大きいのは A 組である.

□

3.4　箱ひげ図

四分位数や四分位範囲を使って, データの分布を図で表してみよう.

データの散らばりのようすを図で表すと, 次のようになる.

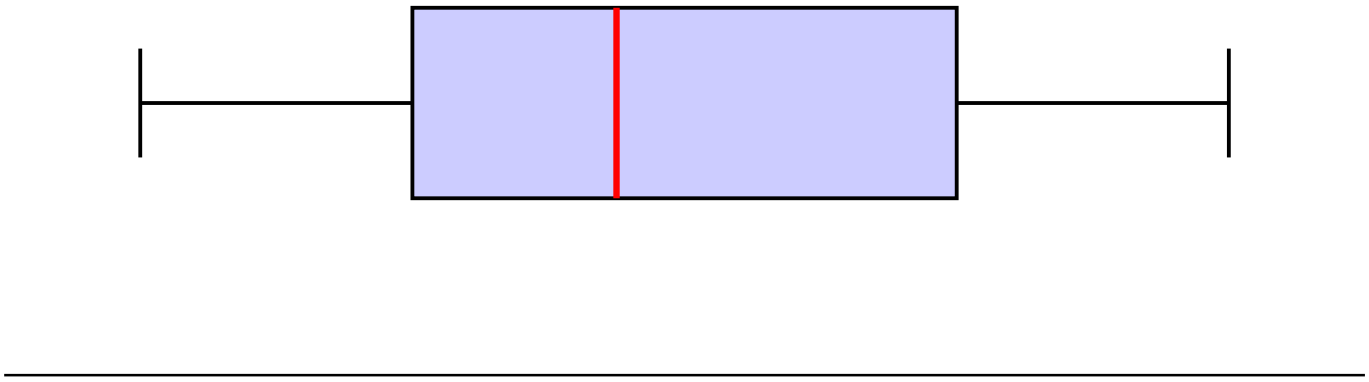

図 3.4.1

Definition 3.4.1　箱ひげ図

図 3.4.1 を **箱ひげ図** という. 箱ひげ図は, データの最小値, 第 1 四分位数, 中央値 (第 2 四分位数), 第 3 四分位数, 最大値を箱とひげで表している. 箱の横の長さは, 四分位範囲を表す.

複数のデータの散らばりのようすを比較する場合は, 箱ひげ図を利用するとよい.

箱ひげ図の「ひげ」は, 最小値や最大値が他のデータの値と大きく離れている場合に, 影響を受けて長くなる.

一方, 「箱」はその影響を受けにくい.

次に, 箱ひげ図とヒストグラムの関係について考えてみよう.

結論から言おう. ヒストグラムの山の位置と, 箱ひげ図の箱の位置がだいたい対応し, ヒストグラムのすそにあたる部分が, 箱ひげ図のひげの部分に対応する. ヒストグラムのすそが左に伸びていれば, 箱ひげ図のひげも左に伸びる.

箱ひげ図では, ヒストグラムほどにはデータの散らばりのようすが表現されないが, 大まかなようすを知ることができる.