

## 四分位数

データを小さい順に並べたとき、それらを四等分する区切りの数を、小さい方から第一四分位数、第二四分位数、第三四分位数と呼ぶ。四分位数は、平均値や標準偏差のように複雑な計算をせずに、データの分布状況を調べるのに便利である。

$n$  個のデータ  $x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n$  の四分位数を求める手順は、次の 2 ステップからなる。

(1) 四分位点  $m_i$  ( $i = 1, 2, 3$ ) を計算する。

(2)  $m_i$  が整数のときは、当然  $m_i$  番目のデータ  $x_{m_i}$  を四分位数  $Q_i$  とする。

$m_i$  が小数になったときは、 $L = \lfloor m_i \rfloor$  番目のデータ  $x_L$  と、 $H = \lceil m_i \rceil$  番目のデータ  $x_H$  から  $Q_i$  を定める。

ただし、 $\lfloor x \rfloor$   $\lceil x \rceil$  はそれぞれ  $x$  の小数部分を切り捨てた値、切り上げた値である。

ステップ 1 の「 $m_i$  の計算式」、ステップ 2 の「 $Q_i$  の定め方」が統一されておらず、下に紹介するように、書物、WEB サイト、統計ソフトによって様々である。当然、得られた値はそれぞれ微妙に異なる。しかし、データの分布状況を調べるという目的に対して悪影響をおよぼすほどの差異はない。特にデータが多数ある場合には全く問題はない。とは言っても、生徒たちが混乱しないように、指導上の注意が必要である。

### 方式 1. A(1)B( $n$ ) を四等分する

(1) 内分点の公式を用いて

$$\begin{aligned} m_1 &= \frac{3 \cdot 1 + 1 \cdot n}{4} & m_2 &= \frac{2 \cdot 1 + 2 \cdot n}{4} & m_3 &= \frac{1 \cdot 1 + 3 \cdot n}{4} \\ &= \frac{n+3}{4} & &= \frac{n+1}{2} & &= \frac{3n+1}{4} \end{aligned}$$



図を見るとわかるように、データの数  $n$  の四等分ではなく、 $n-1$  の四等分になっている。

(2)  $m_i$  が整数でないとき、 $x_L$  と  $x_H$  を小数部に従って比例配分した値を  $Q_i$  とする。

$$Q_i = (H - m_i)x_L + (m_i - L)x_H$$

たとえば、 $m_i = 3.25$  のとき

$$Q_i = 0.75x_3 + 0.25x_4$$

表計算ソフト Excel の関数 QUARTILE はこの方式である。

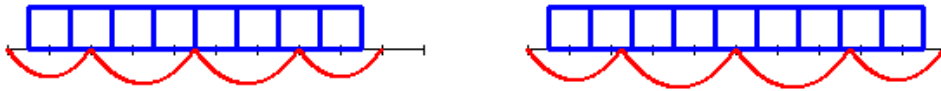
## 方式 2. データの数 $n$ を四等分する

(1)

$$m_1 = \frac{1n}{4} \qquad m_2 = \frac{2n}{4} \qquad m_3 = \frac{3n}{4}$$

中央値がずれてしまう。また、逆順に並べたときの四分位点ともずれてしまう。よって、つぎのように補正する。

$$m_1 = \frac{n}{4} \qquad m_2 = \frac{n+1}{2} \qquad m_3 = n+1 - m_1$$



(2)  $m_i$  が整数でないとき、単に  $x_L$  と  $x_H$  の平均を  $Q_i$  とする。

$$Q_i = \frac{x_L + x_H}{2}$$

## 方式 3. 番号 $i$ を点ではなく $i - 0.5 \sim i + 0.5$ の区間と考える

(1)  $A(0.5)B(n+0.5)$  を四等分する。内分点の公式を用いて

$$\begin{aligned} m_1 &= \frac{3 \cdot 0.5 + 1(n+0.5)}{4} & m_2 &= \frac{2 \cdot 0.5 + 2(n+0.5)}{4} & m_3 &= \frac{1 \cdot 0.5 + 3(n+0.5)}{4} \\ &= \frac{1n+2}{4} & &= \frac{2n+2}{4} & &= \frac{3n+2}{4} \end{aligned}$$



方式 (2) の補正前の値それぞれに  $+0.5$  して補正したものとも考えられる。

(2)  $m_i$  が整数でないとき  $Q_i$  をつぎのように定める。

$$Q_i = x_H \quad (m_i \text{ の小数部が } 0.75, \text{ すなわち } x_H - 0.5 < m_i \text{ のとき})$$

$$Q_i = \frac{x_L + x_H}{2} \quad (m_i \text{ の小数部が } 0.5, \text{ すなわち } x_L + 0.5 = m_i = x_H - 0.5 \text{ のとき})$$

$$Q_i = x_L \quad (m_i \text{ の小数部が } 0.25, \text{ すなわち } m_i < x_L + 0.5 \text{ のとき})$$

最も合理的であるが、実際にはあまり使われていない。

注 以下の方式 4~6 は、まず中央点  $m_2$  と中央値  $Q_2$  を求め、それから四分位点  $m_1, m_3$  と四分位数  $Q_1, Q_3$  を求める方式である。中央値は、次のように定める。

$$m_2 = \frac{1+n}{2}$$

$$Q_2 = \frac{x_L + x_H}{2} \quad (m_2 \text{ が整数のとき } ,x_L = x_H = x_{m_2} \text{ である})$$

#### 方式 4. 中央点以下, 中央点以上の中央値

(1)  $1 \sim m_i$  と  $m_i \sim n$  の中央とする。

$$m_1 = \frac{1 + m_2}{2} = \frac{n + 3}{4} \qquad m_3 = \frac{m_2 + n}{2} = \frac{3n + 1}{4}$$

方式 1 と同じ値になる。分点の公式を使わないのでわかりやすい。

(2)

$$Q_i = \frac{x_L + x_H}{2}$$

#### 方式 5. 中央点以下の整数, 中央点以上の整数の中央値

(1)  $1 \sim [m_2]$  と  $[m_2] \sim n$  の中央とする。

$$m_1 = \frac{1 + [m_2]}{2} \qquad m_3 = \frac{[m_2] + n}{2}$$



(2) 小数部が 0.5 の小数しか現れないので, 単純に平均をとるだけでよい。

$$Q_i = \frac{x_L + x_H}{2}$$

別名ヒンジとも呼ばれる。四分位数とは別物という人もいる。

#### 方式 6. 中央点より小さい整数, 中央点より大きい整数の中央値

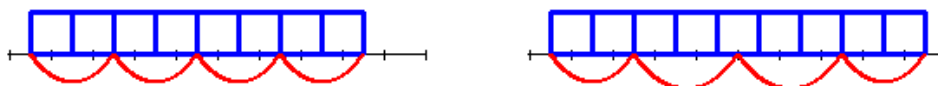
(1) 方式 5 と似ているが, 必ず中央点を除外する。

$m_2$  が整数のとき

$$m_1 = \frac{1 + (m_2 - 1)}{2} \qquad m_3 = \frac{(m_2 + 1) + n}{2}$$

$m_2$  が整数でないとき

$$m_1 = \frac{1 + [m_2]}{2} \qquad m_3 = \frac{[m_2] + n}{2}$$



(2) 小数部が 0.5 の小数しか現れないので, 単純に平均をとるだけでよい。

$$Q_i = \frac{x_L + x_H}{2}$$

## 比較

$m_i$  の図を並べて見ると、微妙な違いがよくわかる。

方式 1



方式 2



方式 3



方式 4



方式 5



方式 6



$n$  を 4 で割った余りで場合分けして,  $Q_1, Q_3$  を表すと次の表のようになる。

$Q_1$

$n$	$4k-3$	$4k-2$	$4k-1$	$4k-0$
方式 1	$x_k$	$\frac{3x_k + x_{k+1}}{4}$	$\frac{x_k + x_{k+1}}{2}$	$\frac{x_k + 3x_{k+1}}{4}$
方式 2	$\frac{x_{k-1} + x_k}{2}$	$\frac{x_{k-1} + x_k}{2}$	$\frac{x_{k-1} + x_k}{2}$	$x_k$
方式 3	$x_k$	$x_k$	$x_k$	$\frac{x_k + x_{k+1}}{2}$
方式 4	$x_k$	$\frac{x_k + x_{k+1}}{2}$	$\frac{x_k + x_{k+1}}{2}$	$\frac{x_k + x_{k+1}}{2}$
方式 5	$x_k$	$x_k$	$\frac{x_k + x_{k+1}}{2}$	$\frac{x_k + x_{k+1}}{2}$
方式 6	$\frac{x_{k-1} + x_k}{2}$	$x_k$	$x_k$	$\frac{x_k + x_{k+1}}{2}$

$Q_3$

$n$	$4k-3$	$4k-2$	$4k-1$	$4k-0$
方式 1	$x_{3k-2}$	$\frac{x_{3k-2} + 3x_{3k-1}}{4}$	$\frac{x_{3k-1} + x_{3k}}{2}$	$\frac{3x_{3k} + x_{3k+1}}{4}$
方式 2	$\frac{x_{3k-2} + x_{3k-1}}{2}$	$\frac{x_{3k-1} + x_{3k}}{2}$	$\frac{x_{3k} + x_{3k+1}}{2}$	$x_{3k+1}$
方式 3	$x_{3k-2}$	$x_{3k-1}$	$x_{3k}$	$\frac{x_{3k} + x_{3k+1}}{2}$
方式 4	$x_{3k-2}$	$\frac{x_{3k-2} + x_{3k-1}}{2}$	$\frac{x_{3k-1} + x_{3k}}{2}$	$\frac{x_{3k} + x_{3k+1}}{2}$
方式 5	$x_{3k-2}$	$x_{3k-1}$	$\frac{x_{3k-1} + x_{3k}}{2}$	$\frac{x_{3k} + x_{3k+1}}{2}$
方式 6	$\frac{x_{3k-2} + x_{3k-1}}{2}$	$x_{3k-1}$	$x_{3k}$	$\frac{x_{3k} + x_{3k+1}}{2}$