

## 4 四分位数

### 4.1 四分位数

四分位数 (quartile) データを等しい大きさの 4 群に分けるような境界値のことを、四分位数または四分位点という。ここでは Tukey の方法を述べる<sup>\*i</sup>。データ  $x_1, x_2, \dots, x_n$  は小さいほうから順に並べられているとする。

1.  $n$  が偶数 ( $n = 2m$ ) のとき、データ全体の中央値を  $Q_2$  とする。

$$Q_2 = \frac{x_m + x_{m+1}}{2}$$

$Q_2$  を境界として下組と上組に分け、

$$\text{下組 } \{x_1, x_2, \dots, x_m\}, \quad \text{上組 } \{x_{m+1}, x_{m+2}, \dots, x_n\}$$

下組の中央値を第 1 四分位数  $Q_1$ 、上組の中央値を第 3 四分位数  $Q_3$  とする。たとえば  $n = 8$  の場合、

$$10, \quad 15, \quad 35, \quad 40, \quad 50, \quad 55, \quad 75, \quad 85$$

の四分位数は、 $Q_1 = 25$ ,  $Q_2 = 45$ ,  $Q_3 = 65$  である。

2.  $n$  が奇数 ( $n = 2m - 1$ ) のとき、データ全体の中央値を  $Q_2$  とする。

$$Q_2 = x_m$$

$Q_2$  を境界として下組と上組に分ける。 $Q_2$  は両方の組に含める。<sup>\*ii</sup>

$$\text{下組 } \{x_1, x_2, \dots, x_m\}, \quad \text{上組 } \{x_m, x_{m+1}, \dots, x_n\}$$

下組の中央値を第 1 四分位数  $Q_1$ 、上組の中央値を第 3 四分位数  $Q_3$  とする。たとえば  $n = 9$  の場合、

$$10, \quad 15, \quad 35, \quad 40, \quad 50, \quad 55, \quad 75, \quad 85, \quad 90$$

の四分位数は、 $Q_1 = 35$ ,  $Q_2 = 50$ ,  $Q_3 = 75$  である。

パーセンタイル (percentile)  $0 < \alpha < 1$  とする。累積相対度数が  $\alpha$  となる値のことを  $100\alpha$  パーセント点またはパーセンタイルといい、 $Q_\alpha$  と表す。25 パーセント点  $Q_{0.25}$  は第 1 四分位数  $Q_1$ 、75 パーセント点  $Q_{0.75}$  は第 3 四分位数  $Q_3$  と考える。ここでは  $\alpha n$  と  $\alpha n + 1$  の中間の順位を用いる方法を述べる。

<sup>\*i</sup>四分位数やパーセンタイルに様々な定義があるが、それらの違いはデータ 1 個分までである。

<sup>\*ii</sup>高校数学の方法では、 $n$  が奇数のとき  $Q_2$  をどちらの組にも含めない。

1.  $j = \alpha n$  が整数のとき,  $x_j$  と  $x_{j+1}$  の平均値を  $Q_\alpha$  とする。

$$Q_\alpha = \frac{x_j + x_{j+1}}{2} \quad (j = \alpha n)$$

たとえば  $n = 8$  の場合,

$$10, 15, 35, 40, 50, 55, 75, 85$$

四分位数の順位  $\alpha n$  が,  $0.25 \times 8 = 2$ ,  $0.75 \times 8 = 6$  だから

$$Q_{0.25} = \frac{x_2 + x_3}{2} = \frac{15 + 35}{2} = 25$$

$$Q_{0.75} = \frac{x_6 + x_7}{2} = \frac{55 + 75}{2} = 65$$

四分位数は,  $Q_1 = 25$ ,  $Q_2 = 45$ ,  $Q_3 = 65$  である。

2.  $\alpha n$  が整数でないとき,  $\alpha n$  の整数部分を  $j$  とし,  $x_{j+1}$  を  $Q_\alpha$  とする。

$$Q_\alpha = x_{j+1} \quad (j = [\alpha n])$$

たとえば  $n = 9$  の場合,

$$10, 15, 35, 40, 50, 55, 75, 85, 90$$

四分位数の順位  $\alpha n$  が,  $0.25 \times 9 = 2.25$ ,  $0.75 \times 9 = 6.75$  だから

$$Q_{0.25} = x_3 = 35$$

$$Q_{0.75} = x_7 = 75$$

四分位数は,  $Q_1 = 35$ ,  $Q_2 = 50$ ,  $Q_3 = 75$  である。

## 4.2 範囲と四分位範囲

**範囲 (range)**  $x_1, x_2, \dots, x_n$  は小さいほうから順に並べられているとする。最大値  $x_n$  から最小値  $x_1$  を引いた値を範囲といい,  $R$  と表す。

$$R = x_n - x_1$$

データに外れ値が含まれているとき, 範囲は外れ値の影響を大きく受ける。

**四分位範囲 (IQR; interquartile range)** 第3四分位数  $Q_3$  から第1四分位数  $Q_1$  を引いた値を四分位範囲といい, IQR と表す。四分位範囲の半分の値を四分位偏差または四分領域といい,  $Q$  と表す。

$$\text{IQR} = Q_3 - Q_1, \quad Q = \frac{Q_3 - Q_1}{2}$$

四分位範囲や四分位偏差は外れ値の影響をほとんど受けない (頑強である)。

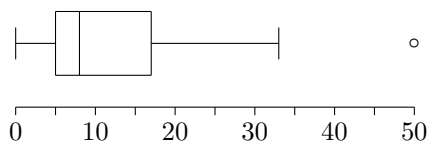
箱ひげ図 (box-and-whisker plot) データの分布を 最小値, 第1四分位数, 中央値, 第3四分位数, 最大値 の組で表すことを 5数要約という<sup>\*iii</sup>。  $Q_1$  から  $Q_3$  までの長さの箱に,  $1.5 \times \text{IQR}$  以内の長さのひげを付けたものを箱ひげ図という。

0 2 3 5 6 7 8 10 12 17 19 33 50

上の 13 個のデータに対して, 四分位数を求める。

$$Q_1 = x_4 = 5, \quad Q_3 = x_{10} = 17$$

5数要約は (0, 5, 8, 17, 50), 箱の長さは  $\text{IQR} = 12$ , ひげの長さは  $1.5 \times \text{IQR} = 18$  以内である。ひげの先端より離れたデータは外れ値として 1 個ずつ書き入れる。



### 4.3 四分位数を用いた頑強な統計量

**Midhinge** Midhinge は第 1 四分位数と第 3 四分位数の midpoint で, 代表値である。

$$\text{Midhinge} = \frac{Q_1 + Q_3}{2}$$

**Quartile coefficient of dispersion** Quartile coefficient of dispersion は, 四分位偏差  $(Q_3 - Q_1)/2$  をミッドヒンジ  $(Q_1 + Q_3)/2$  で割った散布度である。<sup>\*iv</sup> \*iv

$$\text{QCD} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

**Quartile coefficient of skewness** Quartile coefficient of skewness は, 右側の裾の長さから左側の裾の長さを引いた値  $(Q_3 - Q_2) - (Q_2 - Q_1) = Q_1 - 2Q_2 + Q_3$  を四分位範囲  $Q_3 - Q_1$  で割って求める。分布の歪みを表す。

$$\text{QCS} = \frac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1}$$

<sup>\*iii</sup>5数要約や箱ひげ図では, Tukey の方法が用いられることが多い。

<sup>\*iv</sup>総務省統計局や厚生労働省で, ほぼ同じものが使われる。  $\text{QDC} = (Q_3 - Q_1) / (2Q_2)$

## 本稿の参考文献

- 統計学入門（基礎統計学）  
東京大学教養学部統計学教室（編） 東京大学出版会 978-4-13-042065-5
- 統計学  
久保川 達也（著） 東京大学出版会 978-4-13-062921-8
- 共立講座 21世紀の数学 第2版 14 統計  
竹村 彰通（著） 共立出版 978-4-320-01851-8
- 文部科学省検定済教科書 104 数研 数 I 328 高等学校数学 1  
岡部 恒治（ほか著） 数研出版 978-4-410-80203-4
- 統計学序論 改訂版  
山本 義郎（著） 東海大学出版部 978-4-486-02133-9
- 新統計入門  
小寺 平治（著） 裳華房 978-4-7853-1099-8
- Schaum's Outline of Introduction to Probability and Statistics  
Seymour Lipschutz（著） McGraw-Hill Education 978-0-07-176249-6
- [www5e.biglobe.ne.jp/~emm386/statistics/](http://www5e.biglobe.ne.jp/~emm386/statistics/)