

# 四分位数と四分位群：複数定義と用語の区別，その歴史

井口豊\*

\*生物科学研究所，長野県岡谷市

Report of Laboratory of Biology, 17 October 2024

## Quartile values and groups in relation to its multiple definitions and history

Yutaka Iguchi\*

\*Laboratory of Biology, Okaya, Japan

DOI: <https://doi.org/10.5281/zenodo.13889521>

### Abstract

Quartiles represent quartile values or quartile groups. Quartile values represent the three values which divide a population into four equal groups, whereas quartile groups represent the four equal groups divided by the quartile values. Quartiles are calculated by at least nine definitions (e.g. the 1 to 9 types shown by the R software). Japanese high school students learn to use quartiles based on the type 6 for a dataset with an odd number and the type 5 for a dataset with an even number. However, they seem to have few opportunities to learn the history of quantile. Francis Galton may have played an important role in the invention of the idea of quantile.

### 1. はじめに

高校の段階でも学ぶようになった四分位数 (quartile) だが，これには複数の定義があり，高校の教育界で物議を醸したことがある。四分位数の定義（計算の仕方）が，高校授業で学ぶものと，Excelで算出されるもので異なるのである。この問題に関しては，例えば，教科書出版会社「数研」の「教科書の内容に関する Q & A」が参考になる（数研，2014）。この中で，高校数学での四分位数の定義を決めた文部科学省の説明が，「一番簡単な定義である」というのが興味深い。文献情報は，本文末尾に一括掲載してある。

### 2. 高校数学での四分位数の定義とその他の定義

高校数学で定義された四分位数について，例えば，以下のような奇数個，偶数個データ 2 例の四分位数を求める。

- 奇数個 a: 1, 5, 7, 10, 13, 16, 18, 20, 24
- 偶数個 b: 1, 5, 7, 10, 14, 18, 20, 24

具体的な計算については、教科書を見て、各自やってみて欲しい。結果だけ述べると、以下ようになる。

- **a** : 第 1 四分位数 6, 中央値 13, 第 3 四分位数 19
- **b** : 第 1 四分位数 6, 中央値 12, 第 3 四分位数 19

これらのデータを、統計解析ソフト R の `quantile` 関数を使って、9 種類の四分位数を算出してみよう。ここでは、まず自作の `qt` 関数を作り、表形式で、これらの四分位数を表示させる。

さらに、位置の代表値 (measure of location) として、中央値だけでなく、中央ヒンジ (mid-hinge) も計算出力させる。

```
## 9 種類の四分位数を計算・表示する関数 ##
qt<- function(x) {
  table<- matrix(as.numeric(NULL), 9, 3)

  for (type in 1:9) {
    table[type, ]<- quantile(x, c(1/4, 1/2, 3/4), type=type)
  }

  mh<- (table[, 1] + table[, 3])/2
  Type<- 1:9

  table<- cbind(Type, table, mh)

  dimnames(table) <- list(1:9, c("Type", "Q1", "Q2", "Q3", "Midhinge"))
  return(round(table, 3))
}

## データ ##
a<- c(1, 5, 7, 10, 13, 16, 18, 20, 24)
b<- c(1, 5, 7, 10, 14, 18, 20, 24)

## 9 種類の四分位数, 中央値, 中央ヒンジ ##
qt(a)
qt(b)
```

結果は以下の通り。

表 1.9 種類の四分位数, 中央値, 中央ヒンジ

qt (a)				
Type	Q1	Q2	Q3	Midhinge
1	7.000	13.0	18.000	12.5
2	7.000	13.0	18.000	12.5
3	5.000	10.0	18.000	11.5
4	5.500	11.5	17.500	11.5
5	6.500	13.0	18.500	12.5
6	6.000	13.0	19.000	12.5
7	7.000	13.0	18.000	12.5
8	6.333	13.0	18.667	12.5
9	6.375	13.0	18.625	12.5

  

qt (b)				
Type	Q1	Q2	Q3	Midhinge
1	5.000	10	18.000	11.5
2	6.000	12	19.000	12.5
3	5.000	10	18.000	11.5
4	5.000	10	18.000	11.5
5	6.000	12	19.000	12.5
6	5.500	12	19.500	12.5
7	6.500	12	18.500	12.5
8	5.833	12	19.167	12.5
9	5.875	12	19.125	12.5

この結果から, 前述の高校数学での四分位数が, 以下のタイプで計算されていることがわかる。

- 奇数個データ a: タイプ 6
- 偶数個データ b: タイプ 5

高校では, わざわざ, 奇数偶数で異なる定義の計算をしている。しかも前述のとおり, 「一番簡単な定義である」という理由で, このような計算になっている。

### 3. Excel による四分位数の定義

こ Excel 2019 による四分位数を QUARTILE.EXC, QUARTILE.INC, QUARTILE の各関数で求めた結果が以下のとおりである。Excel でも, 関数によって定義が異なることがわかる。ただし, 奇数個・偶数個のような場合分けは無い。

表 2. Excel による四分位数

QUARTILE.EXC			
	Type 6		
	Q1	Q2	Q3
奇数 a	6.0	13.0	19.0
偶数 b	5.5	12.0	19.5

  

QUARTILE.INC			
	Type 7		
	Q1	Q2	Q3
奇数 a	7.0	13.0	18.0
偶数 b	6.5	12.0	18.5

今や中高生でもパソコンを扱う時代である。彼らが将来、統計データを解析するときもパソコンや統計ソフトが頼りとなるであろう。それを見越して、高校生の時から、統計ソフトで四分位数がどのように計算されているか、それを解説するような授業をしてほしい。

四分位数が定義によって異なれば、当然ながら、それを利用した箱ひげ図も異なったものになる。その点も注意が必要である。

#### 4. 四分位数によって分けられたデータの名称

ここまで述べてきた四分位数は、順序化されたデータを4等分する区切りの値 (value) である。一方で、そのようにして4等分されたデータの集団 (group) に対しては、文字通り、四分位群 (quartile group) と呼ばれる。個々の集団は、日本語では、第1、第2、第3、第4四分位群と呼ばれる。論文としては、例えば、深谷ほか (1987) がその例である。その p.755 研究結果を見ると、以下のような記述になっている。

「平均 ± 標準誤差は第1四分位群; 128.4 ± 1.52 mmHg」

それらは英語でも文字通り、the first, second, third, fourth group と呼ばれ、例えば、Foo et al. (2003) は、冒頭の Abstract に以下のように説明していて、理解しやすい。

“The patients were stratified into quartile groups (Q1 to Q4)”

英語の場合、注意すべき点は、quartile 自体に四分位数と四分位群の意味が存在することである。例えば、英語辞書 Oxford Dictionary of English (ODE, Stevenson, 2010) で quartile の最初の定義を見ると、次のように書かれている。

“each of four equal groups into which a population can be divided according to the distribution of values of a particular variable.”

これが、4つに分けられたグループ、つまり「四分位群」を意味することがわかる。

二番目の定義として、いわゆる「四分位数」が載っている。

“each of the three values of the random variable which divide a population into quartiles.”

学術論文で quartile が四分位群の意味で使われている例としては、Kamycheva et al. (2003) がある。その p.103 左段を見ると、以下のような説明となっている。

“Vitamin D intake in both genders was divided into 4 quartiles (2.8, 2.8–5.8, 5.9–10.0, 10.0 g/d)”

つまり、quartiles で quartile groups を表している。

前述の ODE にも載っているように、quartile の語源はラテン語の「四分の一」なので、四分位数よりも、四分位群の定義のほうが、一般の人にとっては自然かもしれない。実際、英語のテレビや新聞などでは、quartile を四分位群の意味で使うほうが多く見られる。私自身も、学生時代、最初に出会った quartile の用例は、四分位群という意味だった記憶がある。

かつて NHK ラジオ第 2 で放送された「実践ビジネス英語」(2018 年 1 月 13 日放送)では、quartile が、統計学の教科書で一般的に使われる「四分位数」の意味ではなく、「四分位群」の意味で解説されていた。もちろん、どちらも正しい用法であるが、文脈によって区別が必要である。四分位数に限らず、一般に分位数を表す英語である「クォンタイル (quantile)」には、等分割する値 (value) の意味と、そのようにして分割された群 (group) の二つの意味がある。英語文献を読んだり、ニュースを聞いたりする際は、その点に注意が必要である。

そのようにして等分割された群に対して、総務省統計局 (2003) は、「分位階級」という用語を用いている。英語で、そのような用語が使われている論文として、例えば、Yemiru, T. et al. (2010) の Table 4 には、Income Quintile class と書かれている。

用語としての quartile の歴史を知るには、Aronson (2001) が役立つ。この論文の最初のページ第 2 段落で、まず quartile の定義に触れられている。そこでは、英語辞書 Oxford English Dictionary (OED) の定義が引用されている。

“each of any set of values of a variate which divide a frequency distribution into equal groups, --- also, any one of the groups so produced”

ここでも、quartile に数 (value) と群 (group) の意味があるとわかる。

それに続いて quartile という単語の歴史が論じられる。これが the late 19th century に現われたということなので、意外と新しい用語なのである。さらに、ある数値の範囲を等分してグループ分けするというアイデアは、Francis Galton によるものなのである。彼は、進化論で有名な Charles Darwin の従兄弟にあたる。ただし OED が、quartile という単語を最初に用いたのは Donald McAlister (1879) であると認定していることも紹介されている。

大学の授業でも、統計量としての四分位数の定義や計算式だけでなく、上記のような語法やその歴史に、簡単で良いので触れてほしいものである。

## 参考文献

Aronson, J. K. (2001) Francis Galton and the invention of terms for quantiles. *Journal of clinical epidemiology*, 54: 1191-1194.

Foo, K. et al. (2003) A single serum glucose measurement predicts adverse outcomes across the whole range of acute coronary syndromes. *Heart*, 89: 512-516.

深谷幸生・大野良之・松本忠雄・荒深美和子（1987）血中鉛量と血圧：某二事業所での鉛検診資料から．日本衛生学雑誌，42(3): 754–761. DOI: 10.1265/jjh.42.754

井口豊（2022）NHK 実践ビジネス英語：統計学用語・四分位 quartile を巡って．  
<https://biolab.sakura.ne.jp/nhk-quartile.html> 2024-10-15 確認．

Kamycheva, E., Joakimsen, R. M. and Jorde, R. (2003) Intakes of calcium and vitamin D predict body mass index in the population of Northern Norway. *The Journal of nutrition* 133: 102-106.  
DOI: 10.1093/jn/133.1.102

McAlister, D. (1879) XIII. The law of the geometric mean. *Proceedings of the Royal Society of London*, 29(196-199), 367-376. DOI: 10.1098/rspl.1879.0061

総務省（2003）用語の説明．IV. 階級区分 <https://www.stat.go.jp/data/zensho/1999/6.htm#04>  
2024-10-15 確認．

Stevenson, A. ed. (2010) *Oxford Dictionary of English (Third ed.)*, Oxford University Press.

数研（2014）教科書の内容に関する Q & A. 数研通信 78.

Yemiru, T. et al. (2010) Forest incomes and poverty alleviation under participatory forest management in the Bale Highlands, Southern Ethiopia. *International Forestry Review*, 12(1): 66-77.