生物科学研究所 研究報告 2024 年 12 月 9 日

四分位偏差とは何か: 変動係数と長野県岡谷市「きなこ石」の話題も含めて

井口豊*
*生物科学研究所，長野県岡谷市

DOI: https://doi.org/10.5281/zenodo.14328888

1. 四分位偏差

四分位偏差（*QD*, quartile deviation）は，統計データのばらつき指標（measure of dispersion）の一つであり，四分位数（quartile）を用いて，以下のように表される。

$$QD = \frac{Q_3 - Q_1}{2}$$

ここで，Q_3 は第 3 四分位数，Q_1 は第 1 四分位数である。要するに，四分位偏差は四分位範囲（quartile deviation）の半分である。この四分位数は四分位点とも呼ばれる。

なお四分位数には複数の定義があり，それによって，四分位偏差の数値も違ってくる。それについては，井口（2024a）を参照してほしい。

中央値を利用したデータのばらつき指標には，他にも中央絶対偏差 (median absolute deviation) がある。

2. 平均と標準偏差に対して，中央値と四分位偏差

中央値と四分位数に対応して，平均と標準偏差の関係を下記のように表すことができる。

第 1 四分位数，中央値，第 3 四分位数
平均 − 標準偏差，平均，平均 + 標準偏差

ここで言う対応関係は，同値であると言う意味ではなく，概念的に対応させると，どうなるか，という意味である。統計学的に厳密な話をするならば，μ と

か， σ とかいった文字を使用するべきだろうが，ここは分かりやすさを優先した記述にする。

略号で簡潔に表すと，次の表 1 のようになる。

表 1. 中央値と四分位数に対する，平均と標準偏差の関係

中央値と四分位数	Q_1	Md	Q_3
平均と標準偏差	$M - SD$	M	$M + SD$

ここで，（平均プラス標準偏差）と（平均マイナス標準偏差）の差を 2 で割ると，標準偏差になる。式で書いたほうが分かりやすい。

$$SD = \frac{(M + SD) - (M - SD)}{2}$$

表 1 の対応関係を見ながら，同様な計算を四分位数におこなうと，それが四分位偏差であることが分かる。

$$QD = \frac{Q_3 - Q_1}{2}$$

3. 四分位偏差は，どんな「ばらつき」を表すのか？

正規分布や一様分布のように，左右対称の形状を持つ確率分布ならば，第 1 四分位数と第 3 四分位数の平均が中央値となる。

$$Md = \frac{Q_1 + Q_3}{2}$$

第 1 四分位数と第 3 四分位数の平均は, 中央ヒンジ（midhinge）と呼ばれる，分布の位置の代表値でもある。

上記の等式は, 左右対称の分布でなければ成り立たないのだが, 後述するように，左右対称分布であるかのように，あるいは，そう見なすような場合が少なくない。

このような不適切な説明が高校教科書にもある，ということで，小林（2013）が批判している。参考文献は, 末尾に一括して挙げたが, 小林（2013）が例えば, p. 66 で取り上げたのは，新 高校の数学 I （数研出版） 132 頁の記述で，四分

位範囲，四分位偏差は，中央値のまわりのデータの散らばり具合を表す値，という説明であった。

小林（2013）は，次のページ p. 67 で，「四分位偏差は中央の 50% のデータがこれこれの範囲に入っている，というだけ」，と批判している。他の教科書も含めて，同様な指摘がされているが，詳しくは，その論文を参照してほしい。

これは,「四分位偏差は中央の **25%**」, または,「四分位**範囲**は中央の 50%」, ということの誤記であろうが,「中央値のまわりの散らばり」と言ってしまう教科書の気持ちも分かる。ここで，左右対称分布ならば，四分位範囲や四分位偏差が，中央値のまわりのデータの散らばりを表すとも言えるからである。

4. 四分位数による相対的なばらつきの指標

データ分布の位置の代表値に対する，相対的なばらつきの大きさの指標として，よく知られたものが変動係数（*CV*, coefficient of variation）であり，標準偏差を平均で割った値として表される。

$$CV = \frac{SD}{M}$$

これまで述べてきたように, これを四分位偏差 *QD* と中央値 *Md* を使って置き換えて, 新たな相対的なばらつき指標 *CQD* を考えると, 以下のようになる。

$$CQD = \frac{QD}{Md}$$

このとき，左右対称な分布を考えると，前述の中央ヒンジを使って，以下のように変形できる。

$$CQD = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}}$$

これを整理すると，以下のようになる。

$$CQD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

この *CQD* は，四分位偏差係数（coefficient of quartile deviation）と呼ばれ，四分位数を使った変動係数と言えるものである。なお，「左右対称な分布を考えると」と述べたが，結果的には，左右非対称な分布であっても適用できる形になっている。しかし，当然であるが，非対称な分布の場合，分母に，中央値を使うか，中央ヒンジを使うかによって，四分位偏差係数の計算結果が異なる。

英語版 Wikipedia では， Quartile coefficient of dispersion という名称になっている（注 1，注釈は末尾に一括）。しかし私としては， dispersion より deviation を使う例のほうが多い気がする。

4. 四分位偏差がどのように使われるか

四分位偏差は，学術論文でもしばしば使われる。例えば，阿部ほか (2014) では， p. 13 右段の結果の記述に，「棒グラフは全被験者の中央値を表しており，エラーバーは四分位偏差を表している」，と書かれている。

内山・山内 (2010) では，p. 202，Table 3 の説明文に， “Median ± Quartile deviation” と書かれている。

Dejima H. et al. (2017) 冒頭の Abstract に， “median ± quartile deviation” と書かれている。

四分位偏差係数についても， Geilert et al. (2020) の p. 8, Fig. 7，ケイ素同位体のグラフで， “median fluid δ^{30}Si value (error bar equals the coefficient of quartile deviation)” つまり，中央値の上下に四分位偏差係数のエラーバーを付けている。

要するに，中央値を挟んだ上下の範囲が四分位範囲，という表現になっている。つまり，第 1 四分位数と第 3 四分位数の平均が中央値となる，という表現なのである。

もちろん，これは左右対称な分布でないと成立しない性質である。しかし，特にそれに触れることなく四分位偏差が使われている。逆に言えば，そのような条件が成り立つ，あるいは，そう見なせるときこそ，簡潔にデータ分布の範囲を示すのに，四分位偏差が有用である，と私は思っている。

ロバスト z スコアを用いた標準化（井口，2024b）も，標準正規分布の平均の両側 50% を四分位範囲に対応させて計算するので，基本的には，左右対称に近い（外れ値があるが），という条件で適用したほうが良いと思われる。

5. 岡谷市の蛇紋岩

統計とは全く関係ないが，前述の Geilert et al. (2020) 論文で研究されているのは，マリアナ前弧の蛇紋岩（serpentinite）である。この蛇紋岩という岩石は，岡谷市の横河川上流域で，中央構造線に相当すると考えられる横河川断層に沿って露出が見られ（吉野，1976），その黄色の模様ゆえに，昔から「きなこ石」として親しまれてきた。この機会に紹介しておこう。

図 1. 岡谷市横河川から産出した蛇紋岩，いわゆる，きなこ石

注

1. Wikipedia. Quartile coefficient of dispersion. https://en.wikipedia.org/wiki/Quartile_coefficient_of_dispersion 2024 年 12 月 9 日確認

参考文献

阿部誠・新沼大樹・吉澤誠・杉田典大・本間経康・山家智之・仁田新一 (2014) 生理的指標を用いた 3 次元映像の生体影響評価における心理的影響の変化. 生体医工学 52(1): 11-17.

Dejima, H., Takahashi, Y., Hato, T., Seto, K., Mizuno, T., Kuroda, H., Sakakura N., Kawamura, M., and Sakao, Y. (2017) Mediastinal pulmonary artery is associated with greater artery diameter and lingular division volume. Scientific reports 7(1); 1-9.

Geilert, S., Grasse, P., Wallmann, K., Liebetrau, V., and Menzies, C. D. (2020) Serpentine alteration as source of high dissolved silicon and elevated δ30Si values to the marine Si cycle. Nature communications 11(1): 1-11.

井口豊 (2024a) 四分位数と四分位群：複数定義と用語の区別，その歴史. 生物科学研究所 研究報告 2024 年 10 月 17 日 Zenodo. https://doi.org/10.5281/zenodo.13889521

井口豊 (2024b) ロバスト z スコア：中央値と四分位数で，非正規分布，外れ値を含む標準化. 生物科学研究所 研究報告 2024 年 12 月 9 日 Zenodo. https://doi.org/10.5281/zenodo.14336057

小林道正 (2013) データ分析における「箱ひげ図」の誤解 ― 高校教科書における多数の誤り ―. 中央大学論集 34: 57-68.

内山敏聡・山内龍男 (2010) 紙における墨のにじみとその評価. 繊維学会誌 66(8): 199-203.

吉野博厚 (1976) 諏訪湖北方および南方の中央構造線: 特に中新世以後の活動について. 地質学論集 13: 61-72.