『データ分析をマスターする 12 のレッスン〔新版〕』

畑農鋭矢・水落正明［著］

補論

発行所　株式会社有斐閣

2022 年 12 月 10 日　初版第 1 刷発行

ISBN 978-4-641-22205-1

©2022, Toshiya Hatano, Masaaki Mizuochi, Printed in Japan

©2022, Toshiya Hatano, Masaaki Mizuochi

コラム：四分位数

本書 93 ページによると，第 1 四分位数，第 2 四分位数，第 3 四分位数は，それぞれ$n/4$，$2n/4(=n/2)$，$3n/4$番目のデータ値のことを指します。ただし，nはデータ数（サイズ）を意味します。ここで，第 2 四分位数は中央値と同じなのですが，84 ページの記述を見ると，nが奇数のとき，中央値は$x_{(n+1)/2}$，またnが偶数のとき，中央値は$\frac{x_{n/2}+x_{(n/2)+1}}{2}$です。どちらが正しいのでしょうか。正確には，$n/4$，$2n/4(=n/2)$，$3n/4$はデータを 4 等分した個数を表しており，第 1 四分位数，第 2 四分位数，第 3 四分位数そのものと一致しません。以下で確認してみますが，四分位数を計算するのは中々面倒な作業であることが分かってもらえると思います。

さて，中央値を考えるときにはデータを 2 等分する必要がありましたので，n が 2 で割り切れる場合（余りが 0，偶数）と割り切れない場合（余りが 1，奇数）とに分けて説明しました。同様に四分位数について考えるときには，データを 4 等分する必要がありますので，nを 4 で割ったときの余りによって分けて説明することになります。いま，$n=12,13,14,15$の 4 ケースを考えることにしましょう。言うまでもありませんが，それぞれ 4 で割ったときの余りが 0,1,2,3 のケースに対応します。まず，余り 0 のケース，すなわち$n=12$のケースを考えましょう。$n/4=3$なので，4 等分された各部には 3 つのデータが収まっているはずです。12 個のデータを 1～12 の数字で表すと，4 等分の境界線はそれぞれ下の青・赤・緑の線のところに位置します。

1,2,3|4,5,6|7,8,9|10,11,12

青線が第 1 四分位数，赤線が第 2 四分位数，緑線が第 3 四分位数に対応することになります。小数で表すと第 1 四分位数が 3.5，第 2 四分位数が 6.5，第 3 四分位数が 9.5 の位置に対応します。つまり，データを 3 個ずつで区切るために，第 1 四分位数は 3 に 0.5 を加えた 3.5 の所に位置し，第 2 四分位数は6（$=3\times2$）に 0.5 を加えた 6.5，第 3 四分位数は9（$=3\times3$）に 0.5 を加えたところに位置するわけです。$n=13,14,15$の各ケースについても同様に考えることができます。$n=13$のケースでは，3.25 個ずつに区切るのですから，同様に 0.5 を加えて四分位数の位置を特定化できます。

【$n=13$（余り 1）のケース】

第 1 四分位数　3.75（$=3.25+0.5$）

第 2 四分位数　7（$=6.5+0.5=3.25\times2+0.5$）

第 3 四分位数　10.25（$= 9.75 + 0.5 = 3.25 \times 3 + 0.5$）

$n = 14$のケースでは 3.5 個ずつに，$n = 15$のケースでは 3.75 個ずつに区切りますので，以下のように四分位数の位置を特定化できます。

【$n = 14$（余り 2）のケース】

第 1 四分位数　4（$= 3.5 + 0.5$）

第 2 四分位数　7.5（$= 7 + 0.5 = 3.5 \times 2 + 0.5$）

第 3 四分位数　11（$= 10.5 + 0.5 = 3.5 \times 3 + 0.5$）

【$n = 15$（余り 3）のケース】

第 1 四分位数　4.25（$= 3.75 + 0.5$）

第 2 四分位数　8（$= 7.5 + 0.5 = 3.75 \times 2 + 0.5$）

第 3 四分位数　11.75（$= 11.25 + 0.5 = 3.75 \times 3 + 0.5$）

ここで，小数点以下.25 や.75 をどのように計算するのかという問題が浮上します。小数点以下の.5 の位置を特定するためには両側のデータの算術平均を用いたことを考えれば，.25 や.75 の位置を特定するためには加重平均を適用すれば計算は可能ですが，その手順は大幅に煩雑になります。そこで，.25 や.75 が出現しない方法がよく使われます。まず中央値で半分に分け，分けたそれぞれをさらに半分に分けて求める方法です。このように求めた四分位数はヒンジと呼ばれますが，中央値を前半にも後半にも含めないところがポイントです。

前と同様に，4 で割ったときの余りが 0,1,2,3 のケースを考えます。この考え方には既に慣れたと思いますので，ここでは最初から一般化して議論します。いま，k を正の整数とすると，nを 4 で割って余りが 0,1,2,3 のケースを，それぞれ$n = 4k$，$n = 4k + 1$，$n = 4k + 2$，$n = 4k + 3$と表すことができます。前に用いた$n = 12,13,14,15$の 4 ケースは$k = 3$の場合に当たるわけです。以下，余りで分けた 4 つのケースについて順に見ていきましょう。

【$n = 4k$】

$n/2 = 2k$となり，2 で割り切れるので（偶数なので），中央値である第 2 四分位数は以下のように表すことができます。

第 2 四分位数　$\frac{x_{n/2} + x_{(n/2)+1}}{2} = \frac{x_{2k} + x_{2k+1}}{2}$（中央値）

次に，第 1 四分位数と第 3 四分位数について考えましょう。また，第 1 四分位数は前半のデータの中央値，第 3 四分位数は後半のデータの中央値です。このデータは中央値を境に$2k$ずつのデータ数に分割できますが，$2k$は2で割り切れますので，前半と後半も偶数のケースの中央値を適用できます。

第 1 四分位数　$\frac{x_k+x_{k+1}}{2}$

第 3 四分位数　$\frac{x_{3k}+x_{3k+1}}{2}$

【$n=4k+1$】

nが 2 で割り切れませんので，中央値である第 2 四分位数は，

第 2 四分位数　$x_{(n+1)/2}=x_{2k+1}$（中央値）

で表されます。

次に，第 1 四分位数と第 3 四分位数について考えましょう。このデータを半分に分割する際に，中央値のデータをいずれにも含めないことにすると，総データ数は$n-1=4k$となり，前半と後半のデータ数は$2k$です。したがって，第 1 四分位数は$n=4k$のケースと同様に，

第 1 四分位数　$\frac{x_k+x_{k+1}}{2}$

です。第 3 四分位数については少し注意してください。第 2 四分位数が存在するため，後半のデータ開始ポイントが 1 つずれていますので，以下のようになります。

第 3 四分位数　$\frac{x_{3k+1}+x_{3k+2}}{2}$

【$n=4k+2$】

2 で割り切れるので（偶数なので），中央値である第 2 四分位数は以下のように表すことができます。

第 2 四分位数　$\frac{x_{n/2}+x_{(n/2)+1}}{2}=\frac{x_{2k+1}+x_{2k+2}}{2}$（中央値）

前半と後半のデータ数は$n/2=2k+1$です。$2k+1$は2で割り切れないので，第 1 四分位数と第 3 四分位数には奇数の場合の中央値を適用します。第 1 四分位数は，

第 1 四分位数　x_{k+1}

となります。第 3 四分位数は，後半がx_{2k+2}から開始することを考えると，

第 3 四分位数　x_{3k+2}

となります。

【$n = 4k + 3$】

2 で割り切れませんので，中央値である第 2 四分位数は，

第 2 四分位数　$x_{(n+1)/2} = x_{2k+2}$（中央値）

で表されます。

前と同様に，前半にも後半にも中央値のデータをいずれにも含めないことにします。総データ数は n-1=4k+2 となり，前半と後半のデータ数は 2k+1 です。2k+1 は 2 で割り切れないので，第 1 四分位数と第 3 四分位数には奇数の場合の中央値を適用します。第 1 四分位数は，

第 1 四分位数　x_{k+1}

となります。第 3 四分位数は，後半がx_{2k+3}から開始することを考えると，

第 3 四分位数　x_{3k+3}

となります。

このようにヒンジの考え方に従うと，.25 や.75 を考える必要がなくなります。なお，一般化された議論が苦手な人は，$k = 3$の場合（$n = 12,13,14,15$）で確認してみるとよいでしょう。四分位数の計算については，以下のサイトの解説も有益です。

「中央値と四分位数の求め方。四分位範囲・四分位偏差とは何か？」

https://atarimae.biz/archives/19162

「四分位数の求め方といろいろな例題」高校数学の美しい物語 https://mathtrain.jp/shibuni

http://math.nakaken88.com/textbook/basic-quartile/